NATIONAL HUMAN GENOME RESEARCH INSTITUTE Division of Intramural Research

Current Topics in Genome Analysis Fall 2006

Week 4: Mining Genomic Sequence Data

Tyra G. Wolfsberg, Ph.D.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES | NATIONAL INSTITUTES OF HEALTH | genome.gov/DIR

Accessing public genome sequence data

UCSC's Genome Browser ("Golden Path")
http://genome.ucsc.edu

NCBI's Map Viewer
http://www.ncbi.nlm.nih.gov/mapview/

Ensembl
http://www.ensembl.org

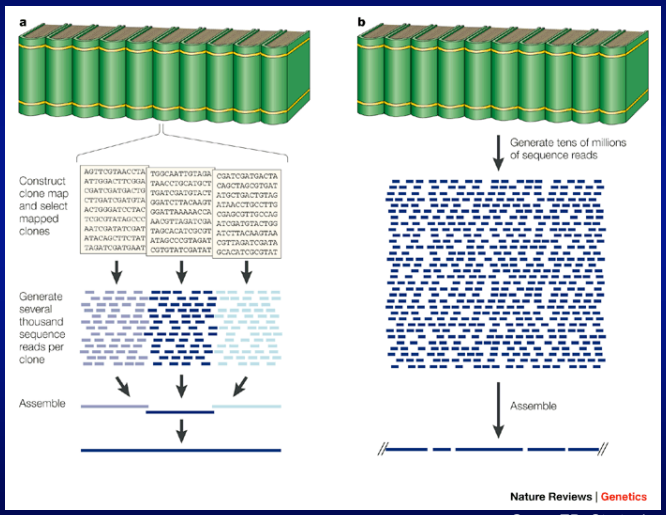
Types of data integrated in genome browsers

- Same starting material for all genome browsers: genomic sequence
- Annotations calculated independently by each genome browser
  - Genes
    - RefSeq mRNAs (non-redundant)
    - GenBank mRNAs (redundant)
    - ESTs
    - Gene predictions
  - SNPs
  - Homologous sequences from other organisms
  - STSs

Overview of genome sequencing strategies

Clone-by-clone shotgun sequencing

Whole-genome shotgun sequencing



Green ED. Strategies for the systematic sequencing of complex genomes. Nat Rev Genet. 2001. 2:573-83.

Genome Sequence Assemblies

- Complex algorithms needed to incorporate all sequence data
- Assemblies updated periodically as new sequence becomes available
  - Mouse and human genomes assembled by NCBI
  - Other genomes assembled by sequencing centers or consortia
- Assemblies not updated concurrently by the three Genome Browsers
  - "Pre-release" assemblies and annotations available at
    - UCSC: *http://genome-test.cse.ucsc.edu/*
    - pre!Ensembl: *http://pre.ensembl.org/*
  - UCSC and Ensembl provide access to older genome assemblies and annotations; NCBI provides access only to old mouse and human data
- IF YOU ARE COMPARING DATA FROM DIFFERENT GENOME BROWSERS, MAKE SURE YOU ARE LOOKING AT THE SAME VERSION OF THE ASSEMBLY

Genome Assembly Versions

	Same assembly?	UCSC	NCBI	Ensembl
Human	Yes	Mar 2006/hg18/Build 36.1	Build 36.1	Build 36
Mouse	YES	Feb 2006/mm8/Build 36	Build 36.1	Build 36
Rat	YES	Nov 2004/rn4/RGSC 3.4	RGSC 3.4	RGSC 3.4
Zebrafish	NO	Mar 2006/danRer4/Zv6	Build 1.1/Zv4	Zv6
Rhesus	YES	Jan 2006/rheMac2/ v.1.0, Mmul_051212	Build 1.1/ v.1.0, Mmul_051212	Mmul_1
Fugu	NO	Aug 2002/ fr1/v3.0	-	Fugu 4.0

NCBI Reference Sequences (RefSeqs)

- Derived from primary GenBank submissions
- Varying levels of validation, additional annotation, and manual curation

NC_123456	Genomic	Mixed	Complete genomic molecules including genomes, chromosomes, organelles, plasmids.
NG_123456	Genomic	Mixed	Incomplete genomic region; supplied to support the NCBI Genome Annotation pipeline.Represents either non-transcribed pseudogenes, or larger regions representing a gene cluster that is difficult to annotate via automatic methods.
NM_123456	mRNA	Mixed	Transcript products; Mature RNA (mRNA) protein-coding transcripts.
NM_123456789	mRNA	Mixed	Transcript products; 9-digit expansion of accession series
NP_123456	Protein	Mixed	Protein products; primarily full-length precursor products but may include some partial proteins and mature peptide products.
NP_123456789	Protein	Curation	Protein products; 9-digit expansion of accession series
NR_123456	RNA	Mixed	Non-coding transcripts including structural RNAs, transcribed pseudogenes, and others
NT_123456	Genomic	Automated	Intermediate genomic assemblies of BAC and/or Whole Genome Shotgun sequence data
NW_123456	Genomic	Automated	Intermediate genomic assemblies of BAC or Whole Genome Shotgun sequence data
NZ_ABCD12345678	Genomic	Automated	A collection of whole genome shotgun sequence data for a project.Accessions are not tracked between releases. The first four characters following the underscore (e.g. 'ABCD') identifies a genome project.
XM_123456	mRNA	Automated	Transcript products; model mRNA provided by the Genome Annotation process; sequence corresponds to the genomic contig.
XP_123456	Protein	Automated	Protein products; model proteins provided by the Genome Annotation process; sequence corresponds to the genomic contig.
XR_123456	RNA	Automated	Transcript products; model non-coding transcripts provided

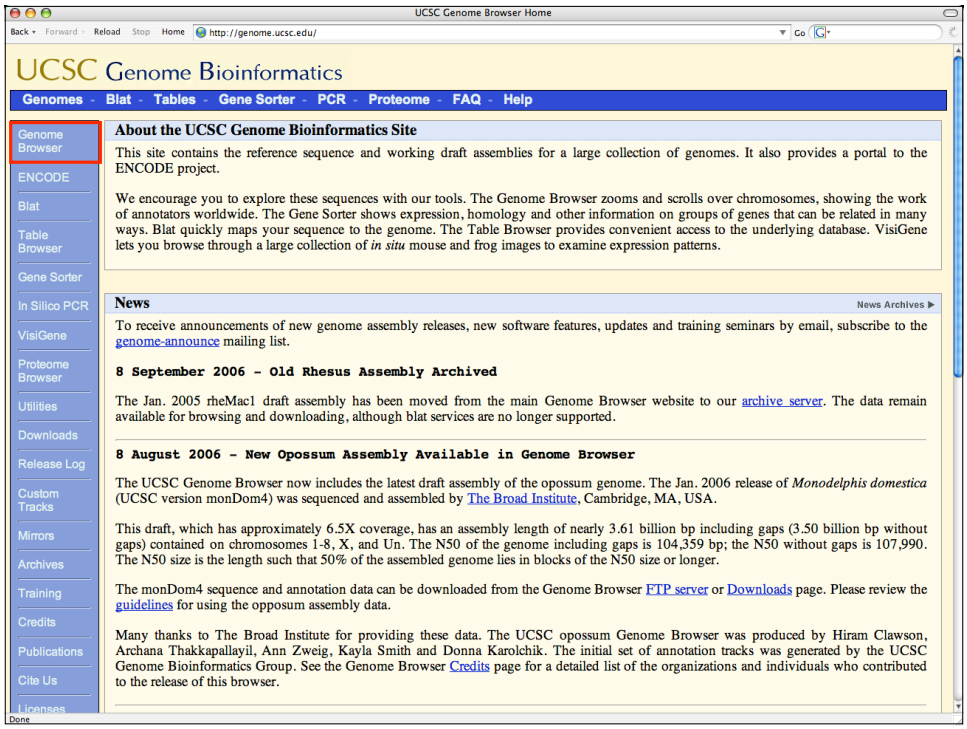
http://www.ncbi.nlm.nih.gov/RefSeq/key.html

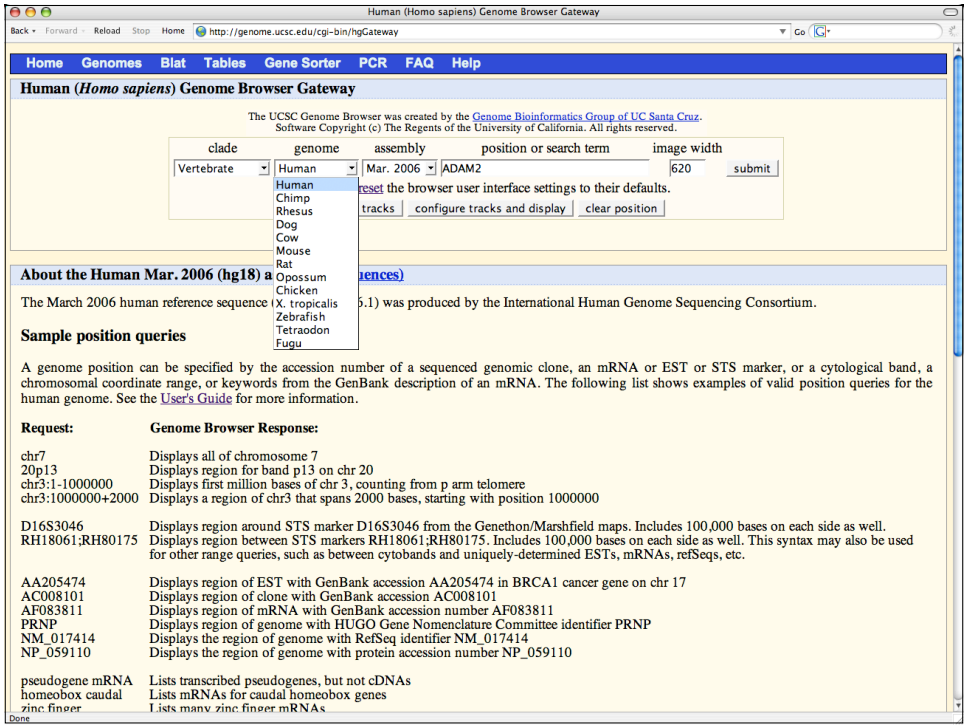
Beta actin mRNA RefSeq

```
LOCUS       NM_001101               1793 bp    mRNA    linear   PRI 17-SEP-2006
DEFINITION  Homo sapiens actin, beta (ACTB), mRNA.
ACCESSION   NM_001101
VERSION     NM_001101.2  GI:5016088
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1793)
  AUTHORS   Pappenberger,G., McCormack,E.A. and Willison,K.R.
  TITLE     Quantitative actin folding reactions using yeast CCT purified via
            an internal tag in the CCT3/gamma subunit
  JOURNAL   J. Mol. Biol. 360 (2), 484-496 (2006)
   PUBMED   16762366

REFERENCE   154 (bases 1 to 1793)
  AUTHORS   Vandekerckhove,J. and Weber,K.
  TITLE     Mammalian cytoplasmic actins are the products of at least two genes
            and differ in primary structure in at least 25 identified positions
            from skeletal muscle actins
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 75 (3), 1106-1110 (1978)
   PUBMED   274701
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from X00351.1 and X63432.1.
            On Jun 8, 1999 this sequence version replaced gi:4501884.

            Summary:  Beta actin is one of six different actin isoforms which
            have been identified.  ACTB is one of the two nonmuscle
            cytoskeletal actins.  Actins are highly conserved proteins that are
            involved in cell motility, structure and integrity.  Alpha actins
            are a major constituent of the contractile apparatus.
            COMPLETENESS: complete on the 3' end.

     CDS             74..1201
                     /gene="ACTB"
                     /gene_component="actin filament; cytoskeleton; TIP60 histone
                     acetyltransferase complex [pmid 10966108]"
                     /gene_function="ATP binding; nucleotide binding; protein
                     binding [pmid 15527767]; structural constituent of
                     cytoskeleton"
                     /note="beta cytoskeletal actin; PS1TP5-binding protein 1"
                     /codon_start=1
                     /product="beta actin"
                     /protein_id="NP_001092.1"
                     /db_xref="GI:4501885"
                     /db_xref="CCDS:CCDS5341.1"
                     /db_xref="GeneID:60"
                     /db_xref="HGNC:132"
                     /db_xref="HPRD:00032"
                     /db_xref="MIM:102630"
                     /translation="MDDDIAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVM
                     VGMGQKDSYVGDEAQSKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHP
                     VLLTEAPLNPKANREKMTQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVT
                     HTVPIYEGYALPHAILRLDLAGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCY
                     VALDFEQEMATAASSSSLEKSYELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHE
                     TTFNSIMKCDVDIRKDLYANTVLSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPE
                     RKYSVWIGGSILASLSTFQQMWISKQEYDESGPSIVHRKCF"

ORIGIN
        1 cgcgtccgcc ccgcgagcac agagcctcgc ctttgccgat ccgccgcccg tccacacccg
       61 ccgccagctc accatggatg atgatatcgc cgcgctcgtc gtcgacaacg gctccggcat
      121 gtgcaaggcc ggcttcgcgg gcgacgatgc cccccgggcc gtcttcccct ccatcgtggg
      181 gcgccccagg caccagggcg tgatggtggg catgggtcag aaggattcct atgtgggcga
      241 cgaggcccag agcaagagag gcatcctcac cctgaagtac cccatcgagc acggcatcgt
```

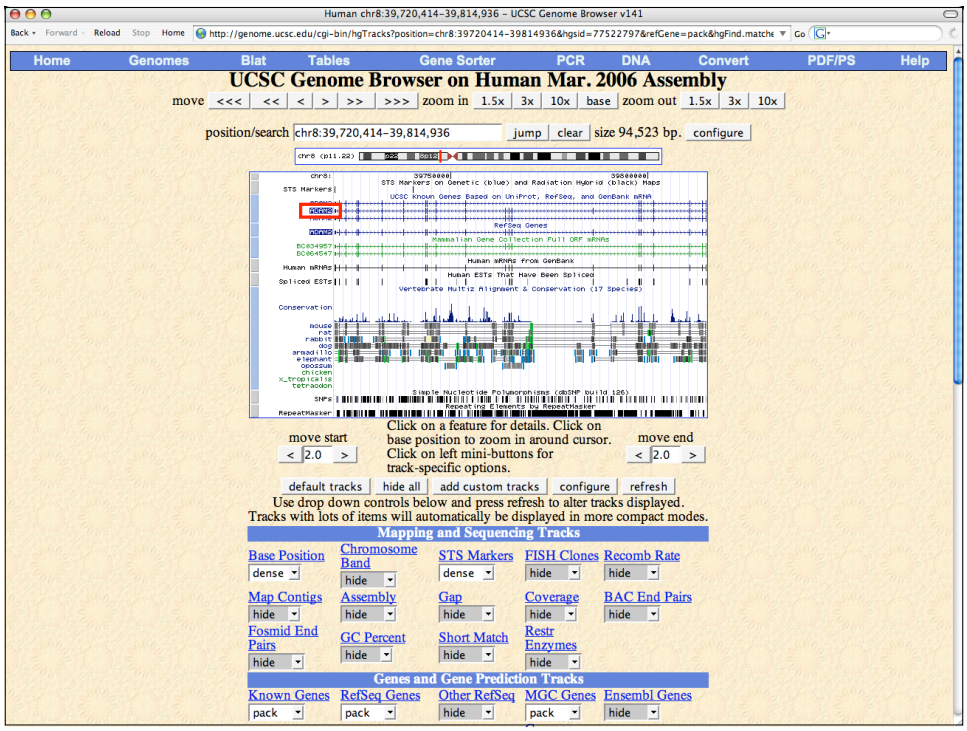
UCSC

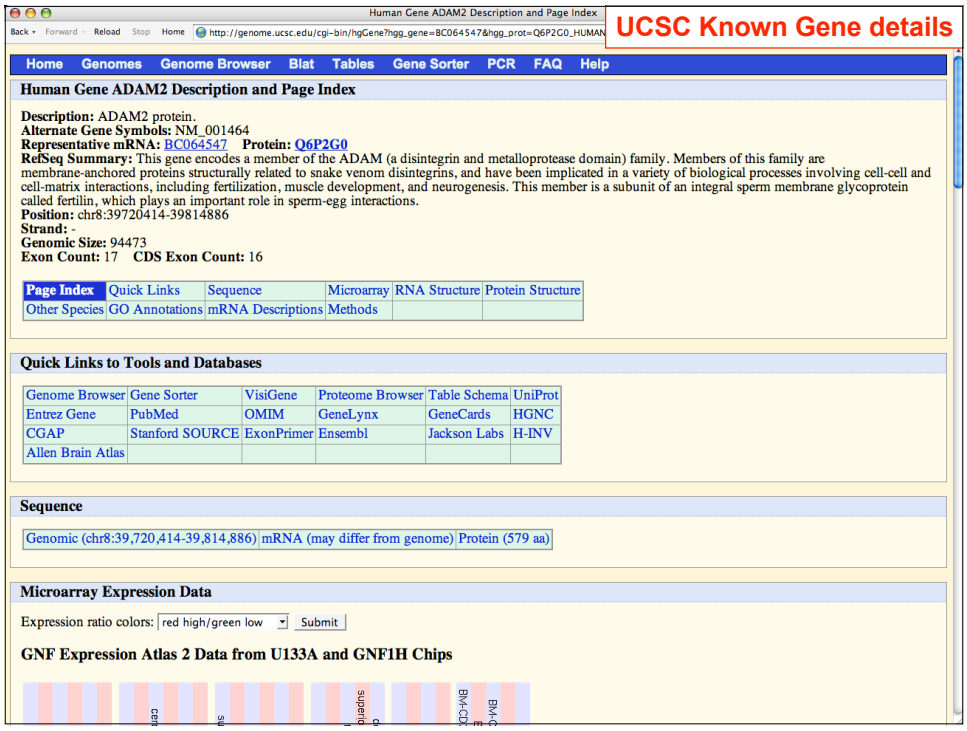
View a region in the genome by querying with a gene symbol

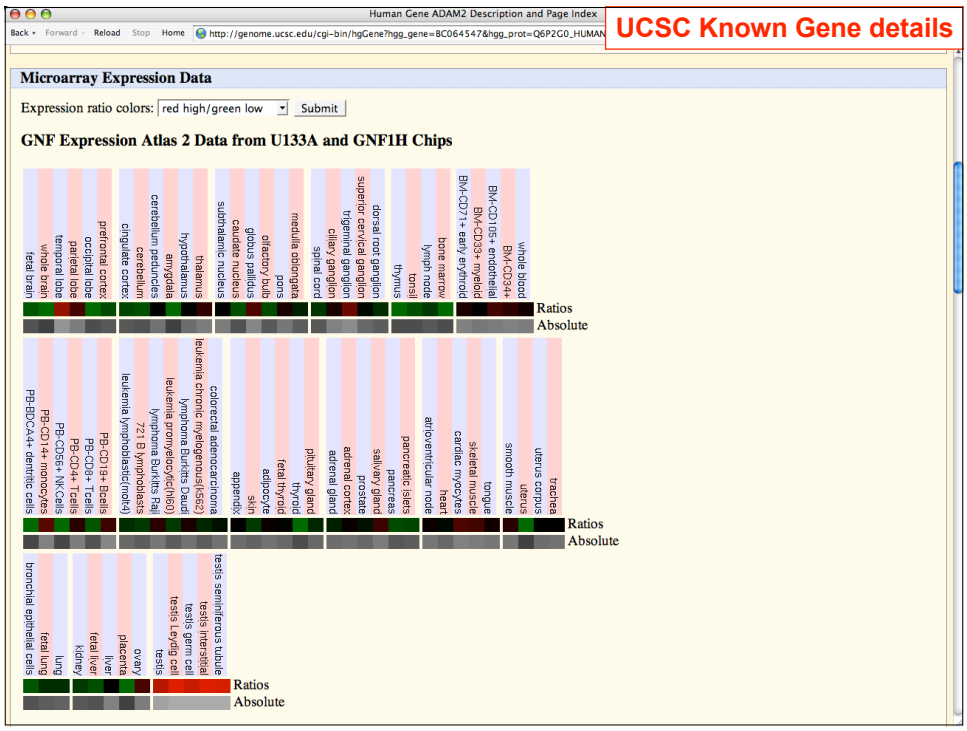


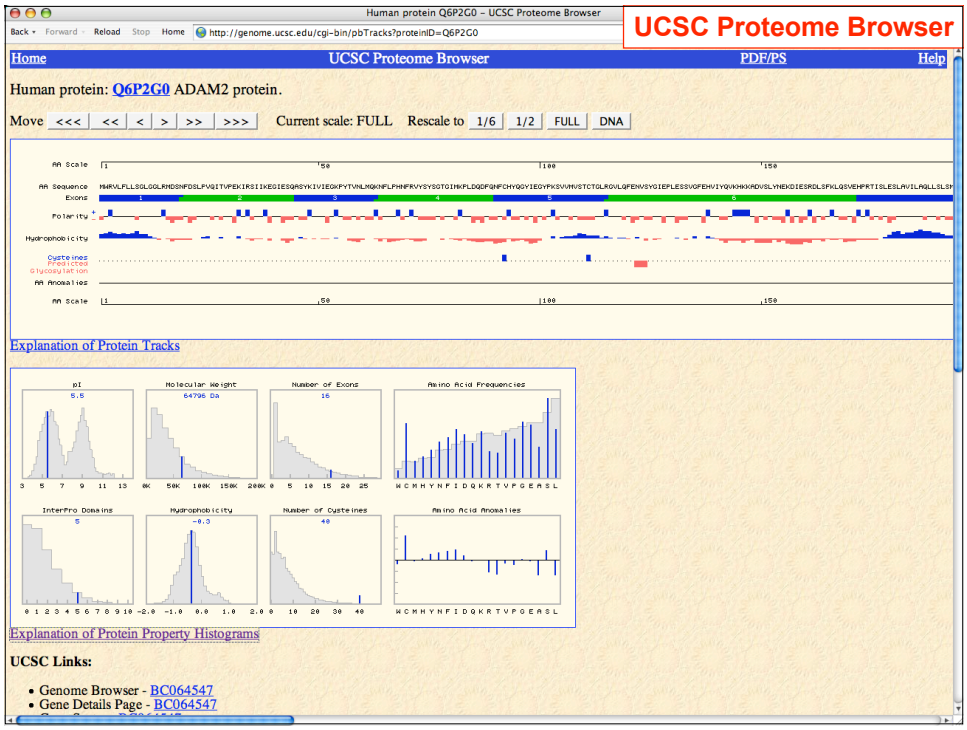

Human (Homo sapiens) Genome Browser Gateway
http://genome.ucsc.edu/cgi-bin/hgGateway
Home Genomes Blat Tables Gene Sorter PCR FAQ Help
Human (Homo sapiens) Genome Browser Gateway
The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.
clade genome assembly position or search term image width
Vertebrate Human Mar. 2006 ADAM2 620 submit
Human
Chimp
Rhesus
Dog
Cow
Mouse
Rat
Opossum
Chicken
X. tropicalis
Zebrafish
Tetraodon
Fugu
reset the browser user interface settings to their defaults.
tracks configure tracks and display clear position
About the Human Mar. 2006 (hg18) a
The March 2006 human reference sequence
Sample position queries
A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, or a cytological band, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the User's Guide for more information.
Request: Genome Browser Response:
chr7 Displays all of chromosome 7
20p13 Displays region for band p13 on chr 20
chr3:1-1000000 Displays first million bases of chr 3, counting from p arm telomere
chr3:1000000+2000 Displays a region of chr3 that spans 2000 bases, starting with position 1000000
D16S3046 Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
RH18061;RH80175 Displays region between STS markers RH18061;RH80175. Includes 100,000 bases on each side as well. This syntax may also be used for other range queries, such as between cytobands and uniquely-determined ESTs, mRNAs, refSeqs, etc.
AA205474 Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC008101 Displays region of clone with GenBank accession AC008101
AF083811 Displays region of mRNA with GenBank accession number AF083811
PRNP Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP
NM_017414 Displays the region of genome with RefSeq identifier NM_017414
NP_059110 Displays the region of genome with protein accession number NP_059110
pseudogene mRNA Lists transcribed pseudogenes, but not cDNAs
homeobox caudal Lists mRNAs for caudal homeobox genes
zinc finger Lists many zinc finger mRNAs
Done

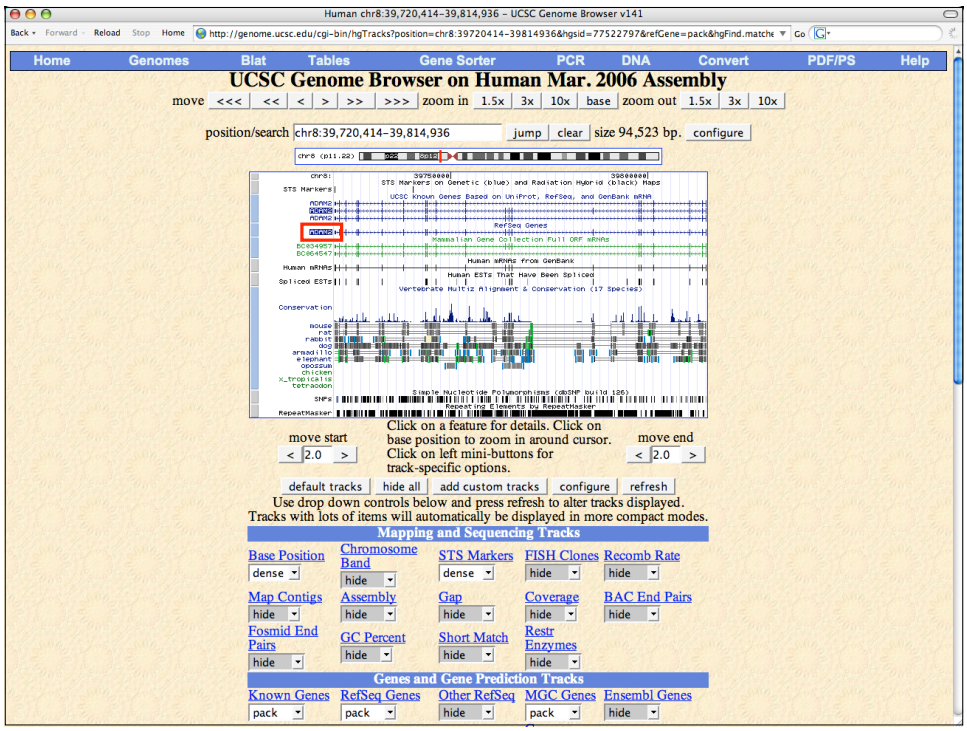

Human ADAM2 - UCSC Genome Browser v141
http://genome.ucsc.edu/cgi-bin/hgTracks?clade=vertebrate&org=Human&db=hg18&position=ADAM2&pix=620&hgsid=7752169
Known Genes
ADAM2 (NM_001464) at chr8:39720414-39814936 - ADAM metallopeptidase domain 2 proprotein
ADAM2 (BC064547) at chr8:39720414-39814886 - ADAM2 protein.
ADAM2 (BC034957) at chr8:39720414-39814885 - ADAM2 protein.
ADAM22 (NM_021723) at chr7:87401638-87664383 - ADAM metallopeptidase domain 22 isoform 1
ADAM22 (NM_021721) at chr7:87401638-87649364 - ADAM metallopeptidase domain 22 isoform 5
ADAM22 (NM_016351) at chr7:87401638-87664383 - ADAM metallopeptidase domain 22 isoform 3
ADAM29 (NM_014269) at chr4:176076134-176135905 - ADAM metallopeptidase domain 29 preproprotein
ADAM28 (NM_014265) at chr8:24207565-24268659 - ADAM metallopeptidase domain 28 isoform 1
ADAM22 (NM_004194) at chr7:87401638-87649364 - ADAM metallopeptidase domain 22 isoform 4
ADAM20 (NM_003814) at chr14:70058832-70071485 - ADAM metallopeptidase domain 20 preproprotein
ADAM21 (NM_003813) at chr14:69993970-69996374 - ADAM metallopeptidase domain 21 preproprotein
ADAM23 (NM_003812) at chr2:207016613-207190922 - ADAM metallopeptidase domain 23 preproprotein
ADAM20 (AF029899) at chr14:70058832-70061255 - ADAM metallopeptidase domain 20
ADAM22 (BC036029) at chr7:87401502-87600049 - ADAM22 protein.
ADAM22 (AF155381) at chr7:87401671-87663791 - Hypothetical protein ADAM22 (Fragment).
ADAM18 (AY358321) at chr8:39561257-39706740 - ADAM metallopeptidase domain 18
RefSeq Genes
click
ADAM2 at chr8:39720414-39814936 - (NM_001464) ADAM metallopeptidase domain 2 proprotein
ADAM20 at chr14:70058832-70071485 - (NM_003814) ADAM metallopeptidase domain 20 preproprotein
ADAM21 at chr14:69993970-69996374 - (NM_003813) ADAM metallopeptidase domain 21 preproprotein
ADAM22 at chr7:87401638-87649364 - (NM_004194) ADAM metallopeptidase domain 22 isoform 4
ADAM22 at chr7:87401638-87649364 - (NM_021721) ADAM metallopeptidase domain 22 isoform 5
ADAM22 at chr7:87401638-87664383 - (NM_021722) ADAM metallopeptidase domain 22 isoform 2
ADAM22 at chr7:87401638-87664383 - (NM_021723) ADAM metallopeptidase domain 22 isoform 1
ADAM22 at chr7:87401638-87664383 - (NM_016351) ADAM metallopeptidase domain 22 isoform 3
ADAM23 at chr2:207016613-207190922 - (NM_003812) ADAM metallopeptidase domain 23 preproprotein
ADAM28 at chr8:24207525-24249555 - (NM_021777) ADAM metallopeptidase domain 28 isoform 3
ADAM28 at chr8:24207525-24268670 - (NM_014265) ADAM metallopeptidase domain 28 isoform 1
ADAM29 at chr4:176076134-176135905 - (NM_014269) ADAM metallopeptidase domain 29 preproprotein
Non-Human RefSeq Genes
ADAM2 at chr8:39723146-39814902 - (NM_213957) fertilin beta
ADAM2 at chr8:39721536-39813888 - (NM_174228) a disintegrin and metalloproteinase domain 2
Adam2 at chr8:39723147-39801569 - (NM_009618) a disintegrin and metalloprotease domain 2
Adam2 at chr8:39723147-39813877 - (NM_020077) a disintegrin and metalloprotease domain 2
Adam21 at chr14:69988547-69996354 - (NM_020330) a disintegrin and metalloprotease domain 21
Adam21 at chr14:69781958-70061197 - (NM_020330) a disintegrin and metalloprotease domain 21
Adam22 at chr7:87401747-87649282 - (NM_001007220) a disintegrin and metalloprotease domain 22
Adam22 at chr7:87401747-87649282 - (NM_001007221) a disintegrin and metalloprotease domain 22
Adam23 at chr2:207016503-207194450 - (NM_011780) a disintegrin and metalloprotease domain 23
Adam23_predicted at chr2:207016759-207190724 - (NM_001029899) a disintegrin and metalloprotease domain 23
Adam24 at chr8:17371373-17373360 - (NM_010086) a disintegrin and metalloproteinase domain 24
Adam24 at chr4:188904672-188906353 - (NM_010086) a disintegrin and metalloproteinase domain 24
Adam24 at chr14:70059197-70060969 - (NM_010086) a disintegrin and metalloproteinase domain 24
Adam25 at chr4:188904411-188906248 - (NM_011781) a disintegrin and metalloprotease domain 25
Adam25 at chr4:176133399-176176696 - (NM_011781) a disintegrin and metalloprotease domain 25
Adam25 at chr14:69994072-69995994 - (NM_011781) a disintegrin and metalloprotease domain 25
Adam25 at chr14:69782108-69783962 - (NM_011781) a disintegrin and metalloprotease domain 25
Adam26a at chr8:17371523-17373287 - (NM_010085) a disintegrin and metalloprotease domain 26

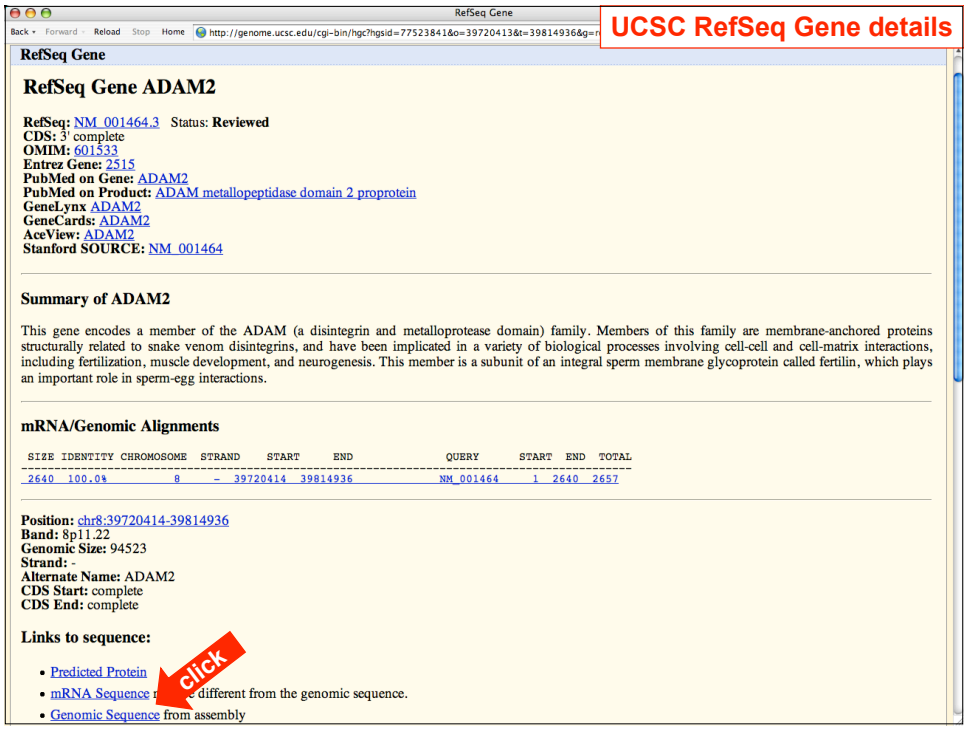

Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v141
Home Genomes Blat Tables Gene Sorter PCR DNA Convert PDF/PS Help
UCSC Genome Browser on Human Mar. 2006 Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position/search chr8:39,720,414-39,814,936 jump clear size 94,523 bp. configure
Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.
move start move end
default tracks hide all add custom tracks configure refresh
Use drop down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
Mapping and Sequencing Tracks
Base Position Chromosome Band STS Markers FISH Clones Recomb Rate
Map Contigs Assembly Gap Coverage BAC End Pairs
Fosmid End Pairs GC Percent Short Match Restr Enzymes
Genes and Gene Prediction Tracks
Known Genes RefSeq Genes Other RefSeq MGC Genes Ensembl Genes

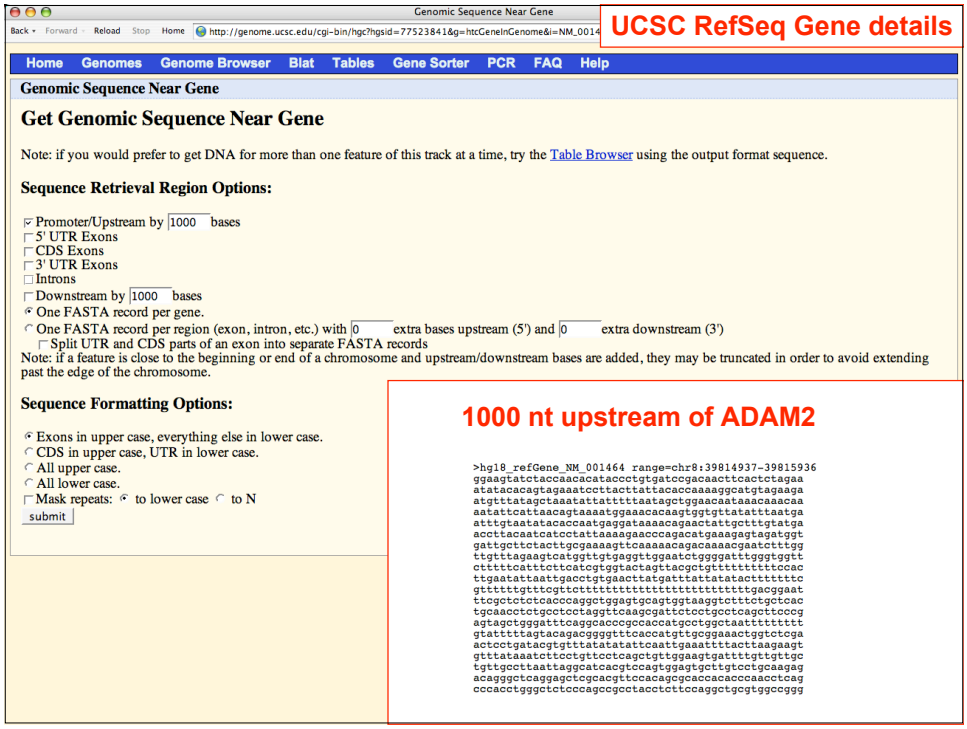

Human Gene ADAM2 Description and Page Index
UCSC Known Gene details
Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help
Human Gene ADAM2 Description and Page Index
Description: ADAM2 protein.
Alternate Gene Symbols: NM_001464
Representative mRNA: BC064547 Protein: Q6P2G0
RefSeq Summary: This gene encodes a member of the ADAM (a disintegrin and metalloprotease domain) family. Members of this family are membrane-anchored proteins structurally related to snake venom disintegrins, and have been implicated in a variety of biological processes involving cell-cell and cell-matrix interactions, including fertilization, muscle development, and neurogenesis. This member is a subunit of an integral sperm membrane glycoprotein called fertilin, which plays an important role in sperm-egg interactions.
Position: chr8:39720414-39814886
Strand: -
Genomic Size: 94473
Exon Count: 17 CDS Exon Count: 16
Page Index Quick Links Sequence Microarray RNA Structure Protein Structure
Other Species GO Annotations mRNA Descriptions Methods
Quick Links to Tools and Databases
Genome Browser Gene Sorter VisiGene Proteome Browser Table Schema UniProt
Entrez Gene PubMed OMIM GeneLynx GeneCards HGNC
CGAP Stanford SOURCE ExonPrimer Ensembl Jackson Labs H-INV
Allen Brain Atlas
Sequence
Genomic (chr8:39,720,414-39,814,886) mRNA (may differ from genome) Protein (579 aa)
Microarray Expression Data
Expression ratio colors: red high/green low Submit
GNF Expression Atlas 2 Data from U133A and GNF1H Chips

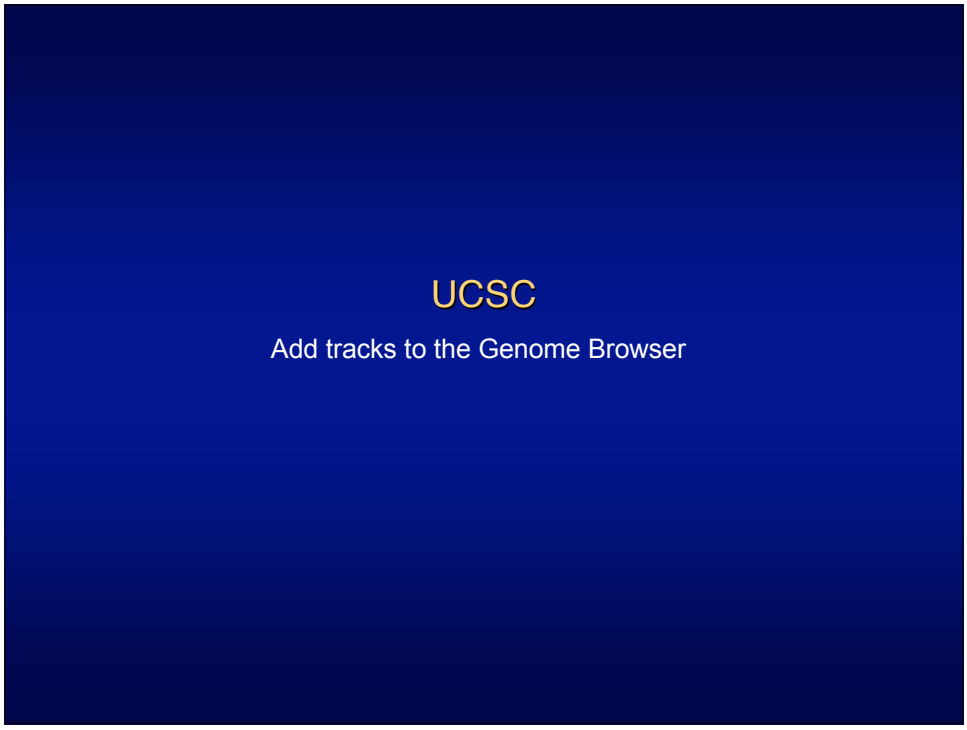

Human Gene ADAM2 Description and Page Index
UCSC Known Gene details
Microarray Expression Data
Expression ratio colors: red high/green low
Submit
GNF Expression Atlas 2 Data from U133A and GNF1H Chips
Ratios
Absolute

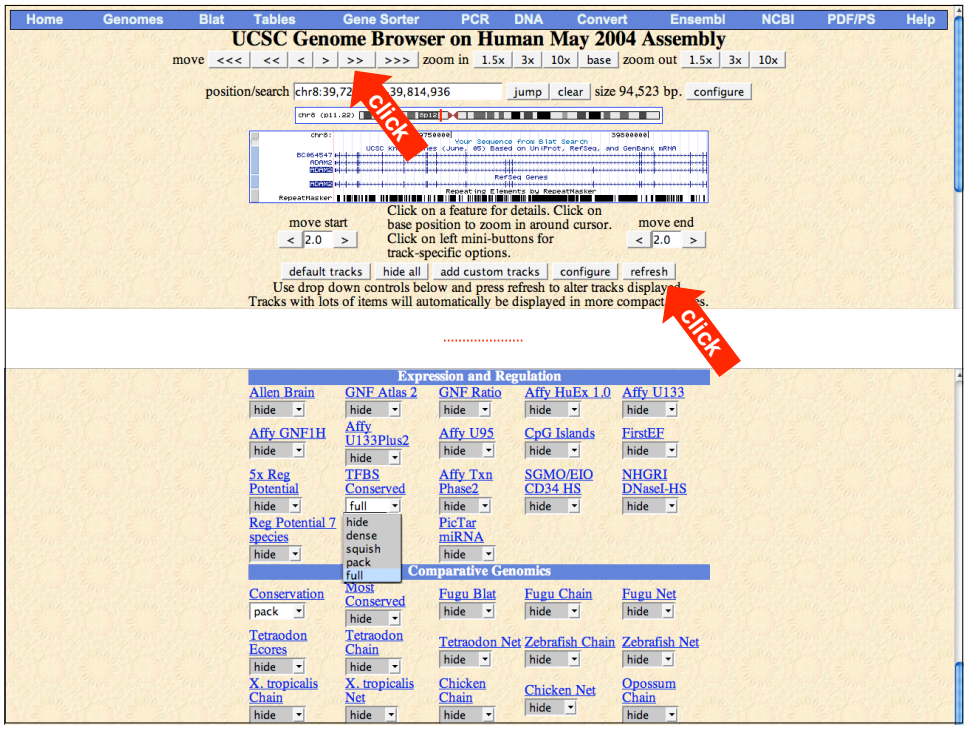

Human protein Q6P2G0 – UCSC Proteome Browser
UCSC Proteome Browser
Home
UCSC Proteome Browser
PDF/PS
Help
Human protein: Q6P2G0 ADAM2 protein.
Move
Current scale: FULL
Rescale to
Explanation of Protein Tracks
Explanation of Protein Property Histograms
UCSC Links:
Genome Browser - BC064547
Gene Details Page - BC064547

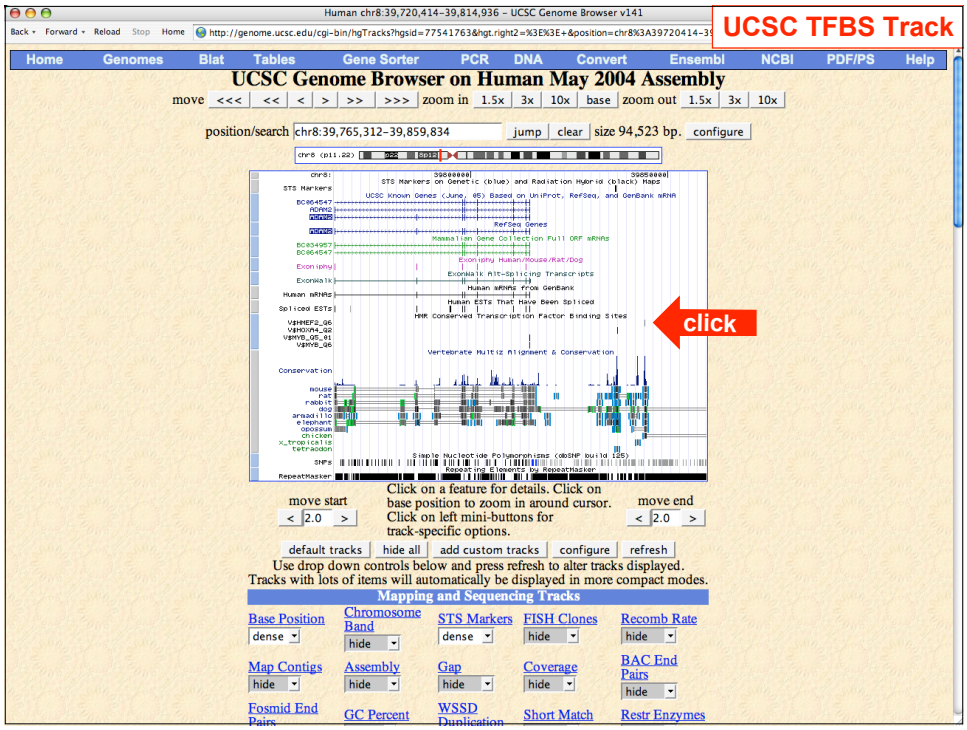

Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v141
Home Genomes Blat Tables Gene Sorter PCR DNA Convert PDF/PS Help
UCSC Genome Browser on Human Mar. 2006 Assembly
position/search chr8:39,720,414-39,814,936 jump clear size 94,523 bp. configure
Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.
move start move end
default tracks hide all add custom tracks configure refresh
Use drop down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
Mapping and Sequencing Tracks
Base Position Chromosome Band STS Markers FISH Clones Recomb Rate
Map Contigs Assembly Gap Coverage BAC End Pairs
Fosmid End Pairs GC Percent Short Match Restr Enzymes
Genes and Gene Prediction Tracks
Known Genes RefSeq Genes Other RefSeq MGC Genes Ensembl Genes

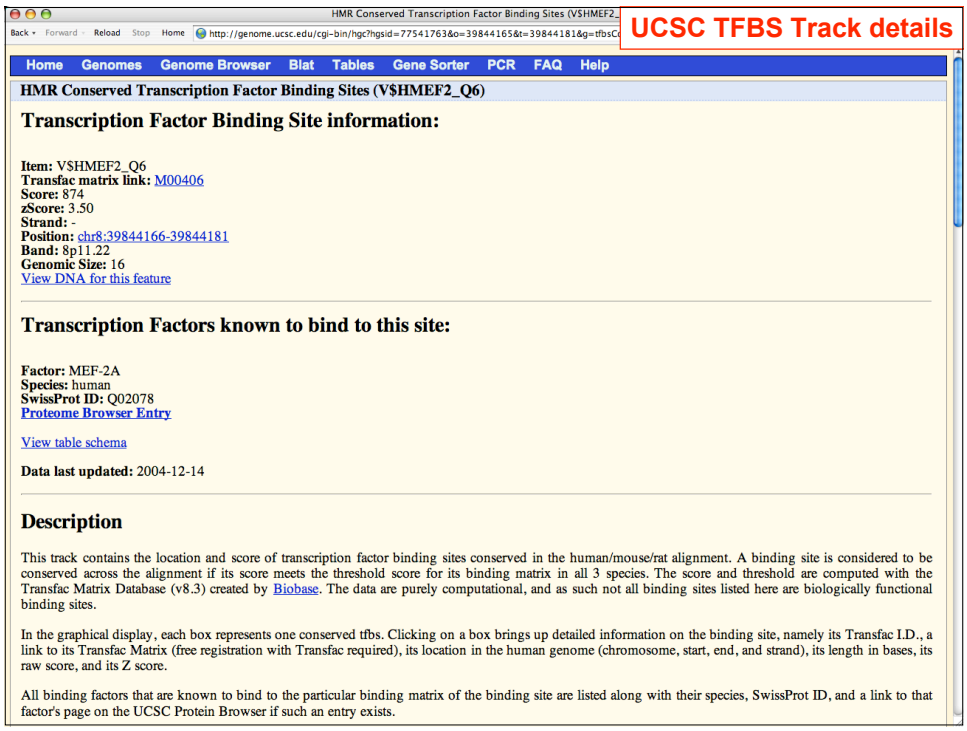

RefSeq Gene
UCSC RefSeq Gene details
RefSeq Gene
RefSeq Gene ADAM2
RefSeq: NM_001464.3 Status: Reviewed
CDS: 3' complete
OMIM: 601533
Entrez Gene: 2515
PubMed on Gene: ADAM2
PubMed on Product: ADAM metallopeptidase domain 2 proprotein
GeneLynx ADAM2
GeneCards: ADAM2
AceView: ADAM2
Stanford SOURCE: NM_001464
Summary of ADAM2
This gene encodes a member of the ADAM (a disintegrin and metalloprotease domain) family. Members of this family are membrane-anchored proteins structurally related to snake venom disintegrins, and have been implicated in a variety of biological processes involving cell-cell and cell-matrix interactions, including fertilization, muscle development, and neurogenesis. This member is a subunit of an integral sperm membrane glycoprotein called fertilin, which plays an important role in sperm-egg interactions.
mRNA/Genomic Alignments
SIZE IDENTITY CHROMOSOME STRAND START END QUERY START END TOTAL
2640 100.0% 8 - 39720414 39814936 NM_001464 1 2640 2657
Position: chr8:39720414-39814936
Band: 8p11.22
Genomic Size: 94523
Strand: -
Alternate Name: ADAM2
CDS Start: complete
CDS End: complete
Links to sequence:
Predicted Protein
mRNA Sequence
different from the genomic sequence.
Genomic Sequence from assembly
click

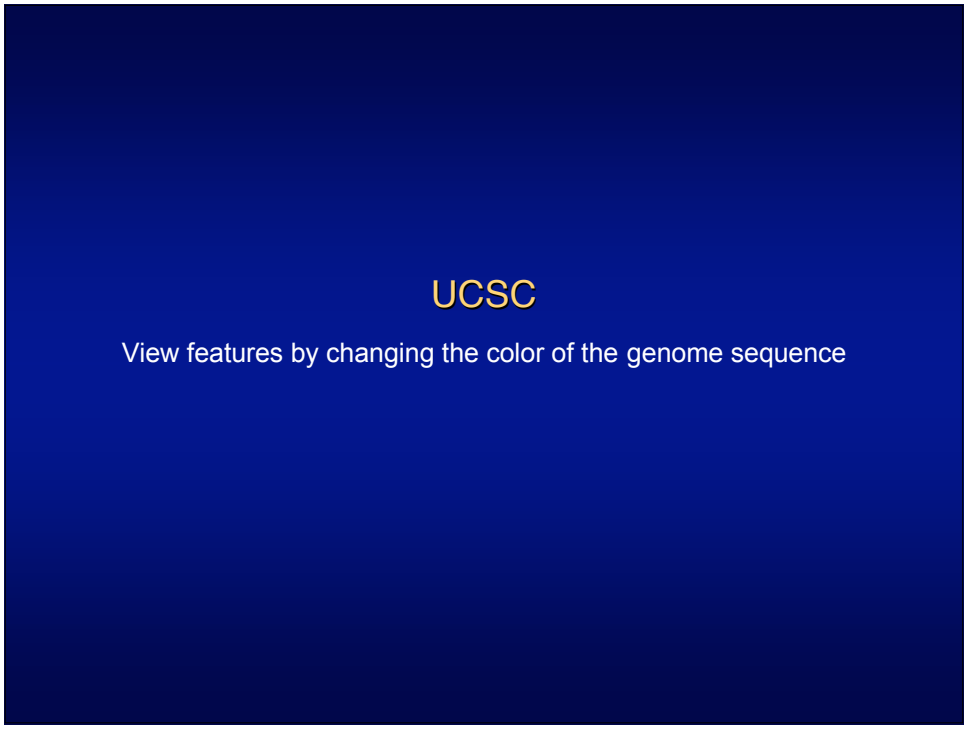

Genomic Sequence Near Gene
UCSC RefSeq Gene details
Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help
Genomic Sequence Near Gene
Get Genomic Sequence Near Gene
Note: if you would prefer to get DNA for more than one feature of this track at a time, try the Table Browser using the output format sequence.
Sequence Retrieval Region Options:
Promoter/Upstream by 1000 bases
5' UTR Exons
CDS Exons
3' UTR Exons
Introns
Downstream by 1000 bases
One FASTA record per gene.
One FASTA record per region (exon, intron, etc.) with 0 extra bases upstream (5') and 0 extra downstream (3')
Split UTR and CDS parts of an exon into separate FASTA records
Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.
Sequence Formatting Options:
Exons in upper case, everything else in lower case.
CDS in upper case, UTR in lower case.
All upper case.
All lower case.
Mask repeats: to lower case to N
submit
1000 nt upstream of ADAM2
>hg18_refGene_NM_001464 range=chr8:39814937-39815936
ggaagtatctaccaacacatacctgtgatccgacaacttcactctagaa
atatacacagtagaaatccttacttattacaccaaaaggcatgtagaaga
atgtttatagctaaatattatttttaatagctggaacaataaacaaacaa
aatattcattaacagtaaaatggaaacacaagtggtgttatatttaatga
atttgtaatatacaccaatgaggataaaacagaactattgctttgtatga
accttacaatcatcctattaaaagaacccagacatgaaagagtagatggt
gattgcttctacttgcgaaaagttcaaaaacagacaaaacgaatctttgg
ttgtttagaagtcatggttgtgaggttggaatctggggatttgggtggtt
ctttttcatttcttcatcgtggtactagttacgctgtttttttttttccac
ttgaatattaattgacctgtgacttatgatttatttatatactttttttc
gttttttgtttcgttctttttttttttttttttttttttttgacggaat
ttcgctctctcacccaggctggagtgcagtggtaaggtctttctgctcac
tgcaacctctgcctcctaggttcaagcgattctcctgcctcagcttcccg
agtagctgggatttcaggcacccgccaccatgcctggctaatttttttt
gtattttagtacagacggggttcaccatgttgcggaaactggtctcga
actcctgatacgtgtttatatatattcaattgaaatttacttaagaagt
gtttataaatcttcctgttcctcagctgttggaagtgattttgttgttgc
tgttgccttaattaggcatcacgtccagtggagtgcttgtcctgcaagag
acagggctcaggagctcgcacgttccacagcgcaccacacccaacctcag
cccacctgggctctcccagcgcctacctcttccaggctgcgtggccggg

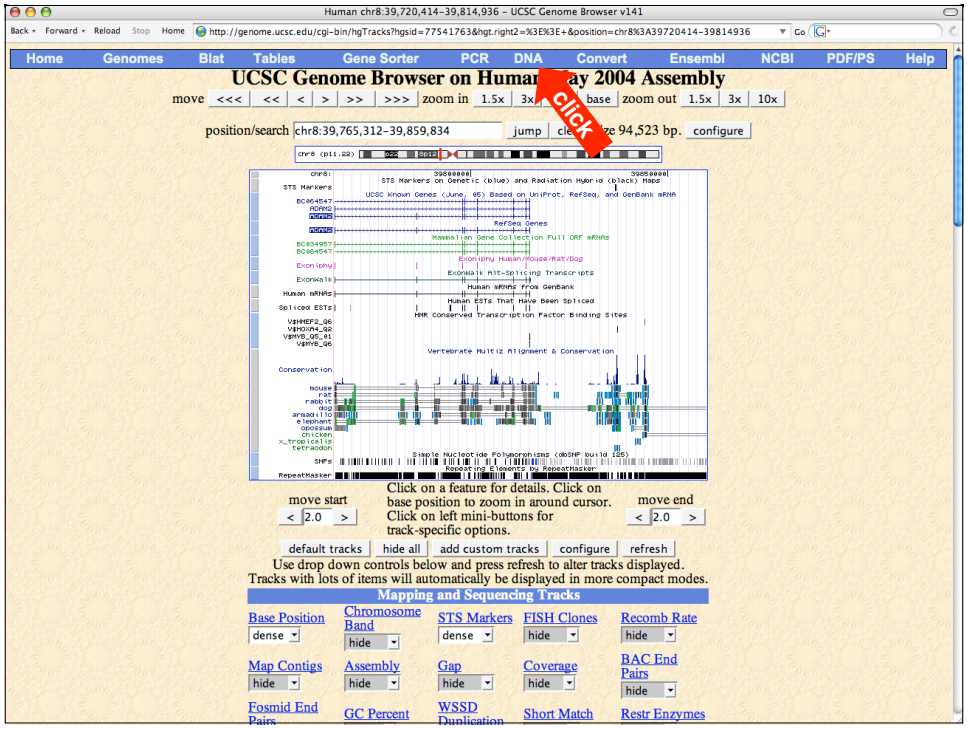

UCSC
Add tracks to the Genome Browser

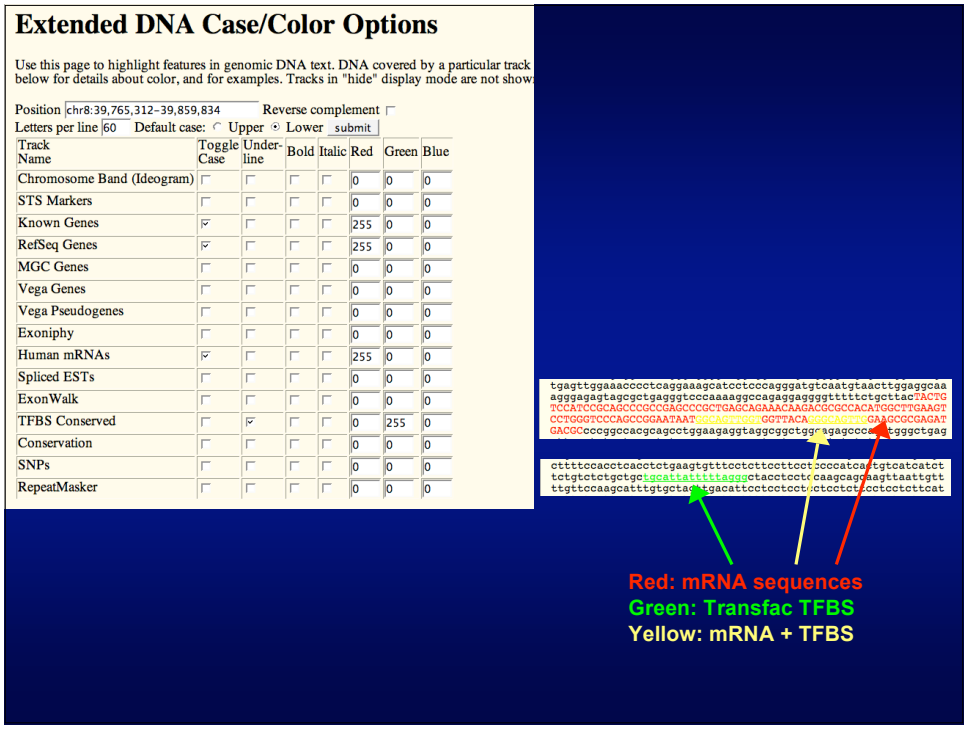

Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Help
UCSC Genome Browser on Human May 2004 Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position/search chr8:39,72 39,814,936 jump clear size 94,523 bp. configure
click
move start move end
Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.
default tracks hide all add custom tracks configure refresh
Use drop down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
click
Expression and Regulation
Allen Brain
GNF Atlas 2
GNF Ratio
Affy HuEx 1.0
Affy U133
Affy GNF1H
Affy U133Plus2
Affy U95
CpG Islands
FirstEF
5x Reg Potential
TFBS Conserved
Affy Txn Phase2
SGMO/EIO CD34 HS
NHGRI DNaseI-HS
Reg Potential 7 species
PicTar miRNA
hide dense squish pack full
Comparative Genomics
Conservation
Most Conserved
Fugu Blat
Fugu Chain
Fugu Net
Tetraodon Ecores
Tetraodon Chain
Tetraodon Net
Zebrafish Chain
Zebrafish Net
X. tropicalis Chain
X. tropicalis Net
Chicken Chain
Chicken Net
Opossum Chain


UCSC TFBS Track
Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v141
Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Help
UCSC Genome Browser on Human May 2004 Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position/search chr8:39,765,312-39,859,834 jump clear size 94,523 bp. configure
click
Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.
default tracks hide all add custom tracks configure refresh
Use drop down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
Mapping and Sequencing Tracks
Base Position
Chromosome Band
STS Markers
FISH Clones
Recomb Rate
Map Contigs
Assembly
Gap
Coverage
BAC End Pairs
Fosmid End Pairs
GC Percent
WSSD Duplication
Short Match
Restr Enzymes


UCSC TFBS Track details
HMR Conserved Transcription Factor Binding Sites (V$HMEF2_
Back Forward Reload Stop Home
http://genome.ucsc.edu/cgi-bin/hgc?hgsid=77541763&o=39844165&t=39844181&g=tfbsC
Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help
HMR Conserved Transcription Factor Binding Sites (V$HMEF2_Q6)
Transcription Factor Binding Site information:
Item: V$HMEF2_Q6
Transfac matrix link: M00406
Score: 874
zScore: 3.50
Strand: -
Position: chr8:39844166-39844181
Band: 8p11.22
Genomic Size: 16
View DNA for this feature
Transcription Factors known to bind to this site:
Factor: MEF-2A
Species: human
SwissProt ID: Q02078
Proteome Browser Entry
View table schema
Data last updated: 2004-12-14
Description
This track contains the location and score of transcription factor binding sites conserved in the human/mouse/rat alignment. A binding site is considered to be conserved across the alignment if its score meets the threshold score for its binding matrix in all 3 species. The score and threshold are computed with the Transfac Matrix Database (v8.3) created by Biobase. The data are purely computational, and as such not all binding sites listed here are biologically functional binding sites.
In the graphical display, each box represents one conserved tfbs. Clicking on a box brings up detailed information on the binding site, namely its Transfac I.D., a link to its Transfac Matrix (free registration with Transfac required), its location in the human genome (chromosome, start, end, and strand), its length in bases, its raw score, and its Z score.
All binding factors that are known to bind to the particular binding matrix of the binding site are listed along with their species, SwissProt ID, and a link to that factor's page on the UCSC Protein Browser if such an entry exists.

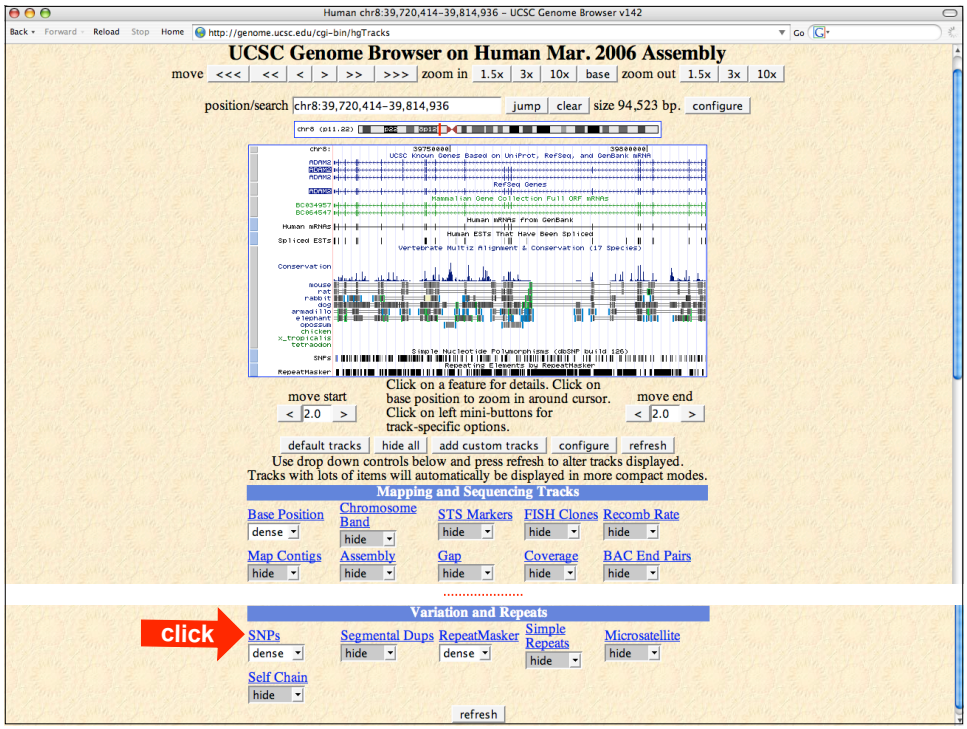

UCSC
View features by changing the color of the genome sequence

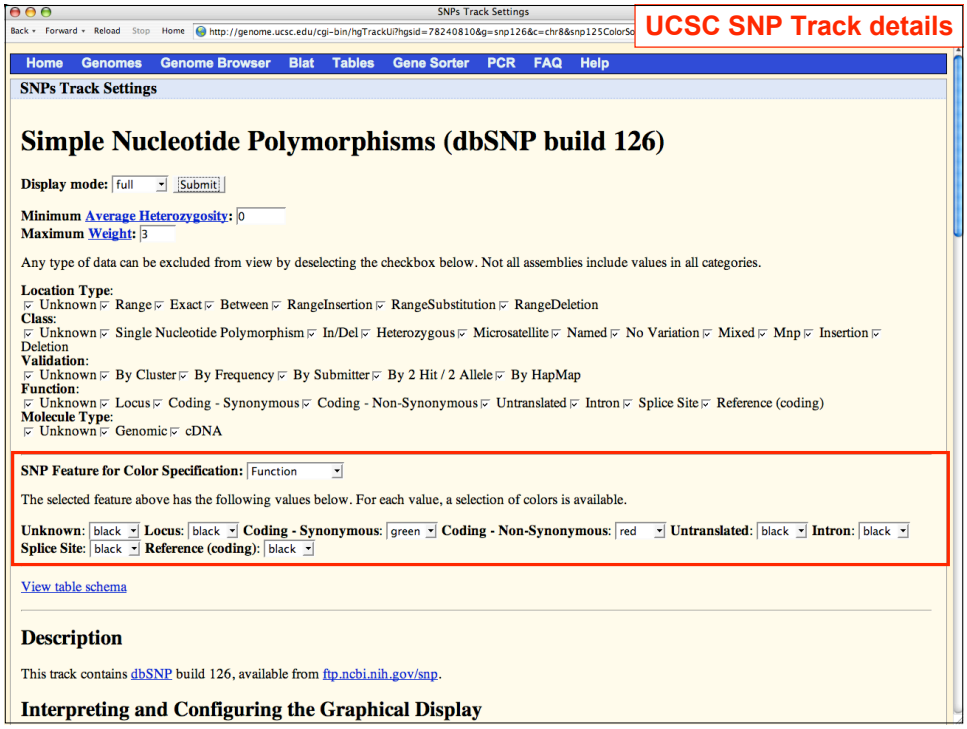

Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v141
Home
Genomes
Blat
Tables
Gene Sorter
PCR
DNA
Convert
Ensembl
NCBI
PDF/PS
Help
click
position/search chr8:39,765,312-39,859,834
jump
configure
Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.
move start
move end
default tracks
hide all
add custom tracks
configure
refresh
Use drop down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
Mapping and Sequencing Tracks
Base Position
Chromosome Band
STS Markers
FISH Clones
Recomb Rate
Map Contigs
Assembly
Gap
Coverage
BAC End Pairs
Fosmid End Pairs
GC Percent
WSSD Duplication
Short Match
Restr Enzymes

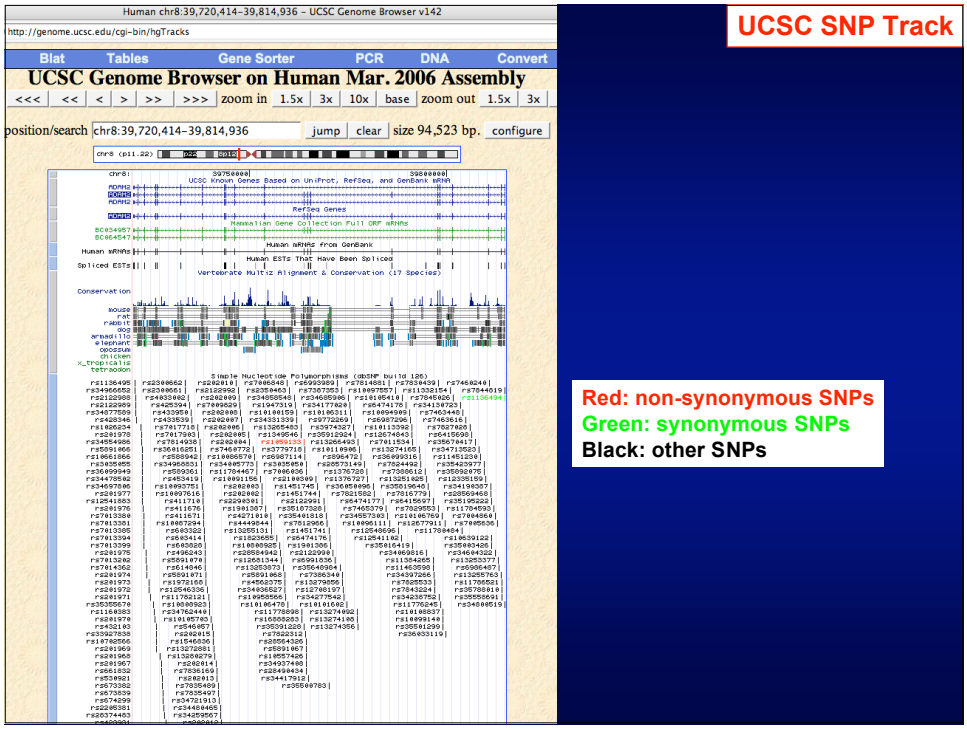

Extended DNA Case/Color Options
Use this page to highlight features in genomic DNA text. DNA covered by a particular track
below for details about color, and for examples. Tracks in "hide" display mode are not show
Position chr8:39,765,312-39,859,834
Reverse complement
Letters per line 60
Default case: Upper Lower
submit
Track Name
Toggle Case
Under-line
Bold
Italic
Red
Green
Blue
Chromosome Band (Ideogram)
STS Markers
Known Genes
RefSeq Genes
MGC Genes
Vega Genes
Vega Pseudogenes
Exoniphy
Human mRNAs
Spliced ESTs
ExonWalk
TFBS Conserved
Conservation
SNPs
RepeatMasker
Red: mRNA sequences
Green: Transfac TFBS
Yellow: mRNA + TFBS

UCSC

Change the color of items in a track

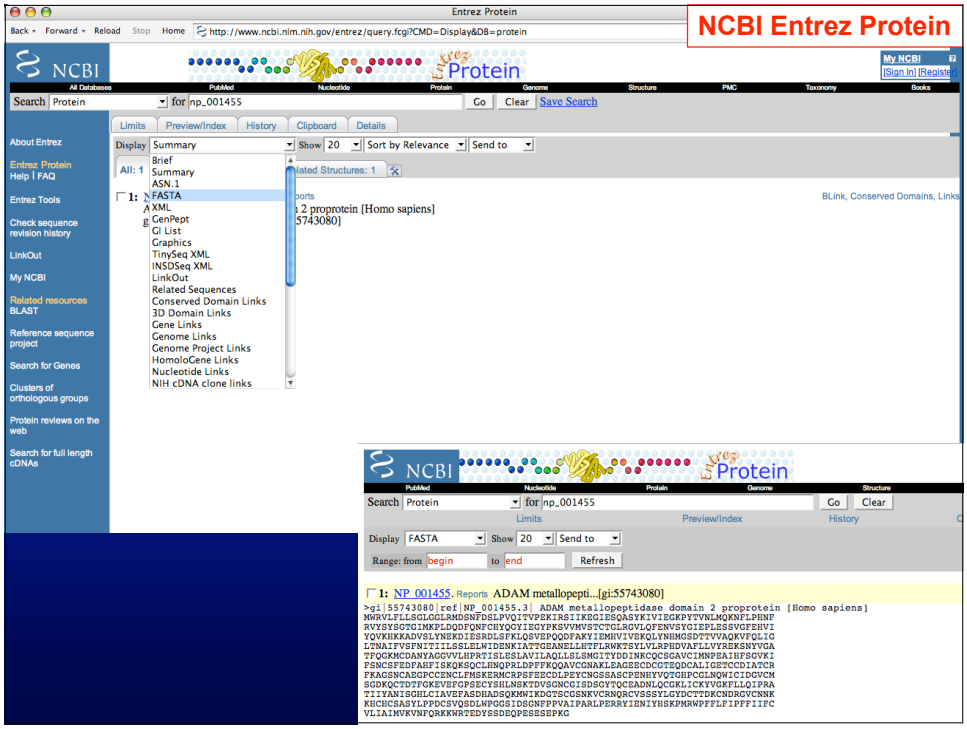


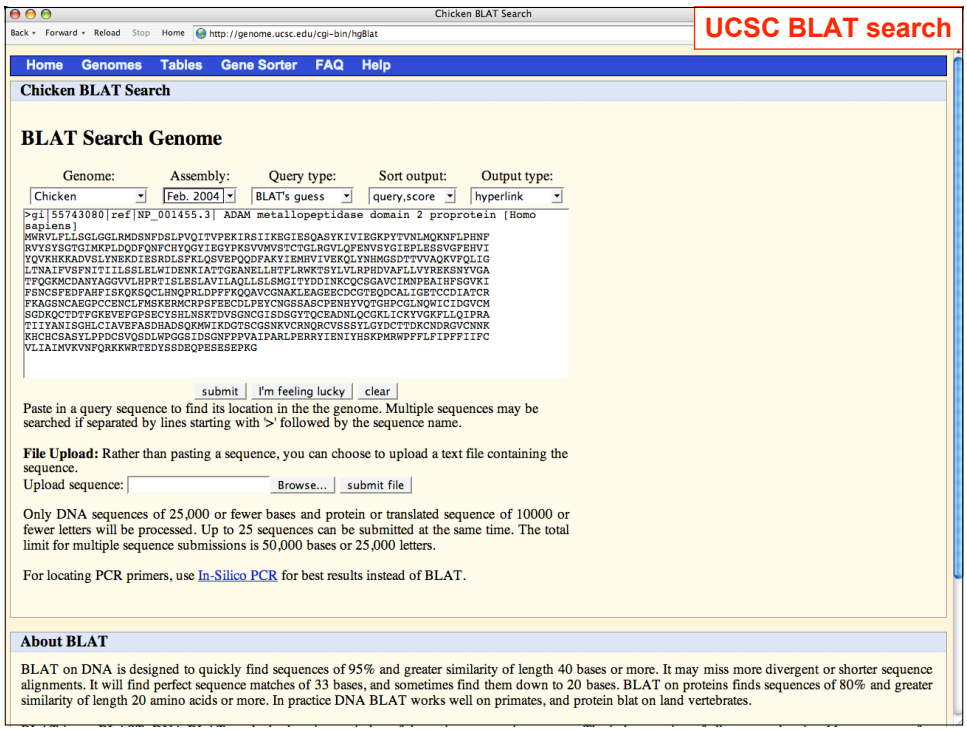

UCSC SNP Track details
SNPs Track Settings
Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help
SNPs Track Settings
Simple Nucleotide Polymorphisms (dbSNP build 126)
Display mode: full Submit
Minimum Average Heterozygosity: 0
Maximum Weight: 3
Any type of data can be excluded from view by deselecting the checkbox below. Not all assemblies include values in all categories.
Location Type:
Unknown Range Exact Between RangeInsertion RangeSubstitution RangeDeletion
Class:
Unknown Single Nucleotide Polymorphism In/Del Heterozygous Microsatellite Named No Variation Mixed Mnp Insertion Deletion
Validation:
Unknown By Cluster By Frequency By Submitter By 2 Hit / 2 Allele By HapMap
Function:
Unknown Locus Coding - Synonymous Coding - Non-Synonymous Untranslated Intron Splice Site Reference (coding)
Molecule Type:
Unknown Genomic cDNA
SNP Feature for Color Specification: Function
The selected feature above has the following values below. For each value, a selection of colors is available.
Unknown: black Locus: black Coding - Synonymous: green Coding - Non-Synonymous: red Untranslated: black Intron: black
Splice Site: black Reference (coding): black
View table schema
Description
This track contains dbSNP build 126, available from ftp.ncbi.nih.gov/snp.
Interpreting and Configuring the Graphical Display

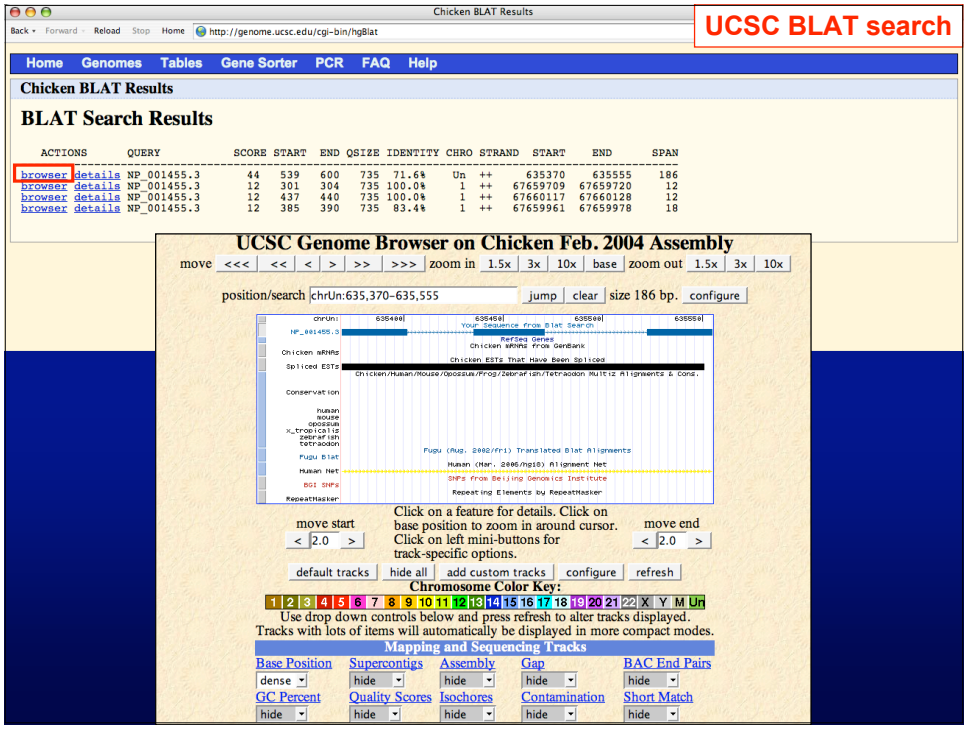

UCSC SNP Track
Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v142
http://genome.ucsc.edu/cgi-bin/hgTracks
Blat Tables Gene Sorter PCR DNA Convert
UCSC Genome Browser on Human Mar. 2006 Assembly
zoom in 1.5x 3x 10x base zoom out 1.5x 3x
position/search chr8:39,720,414-39,814,936 jump clear size 94,523 bp. configure
Red: non-synonymous SNPs
Green: synonymous SNPs
Black: other SNPs

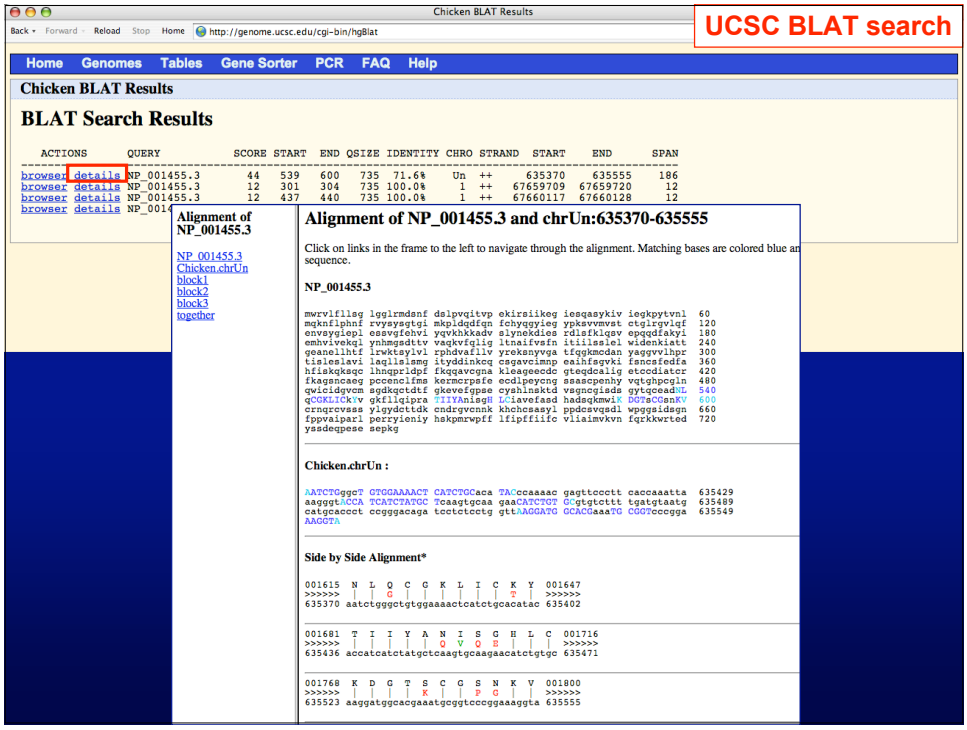
UCSC

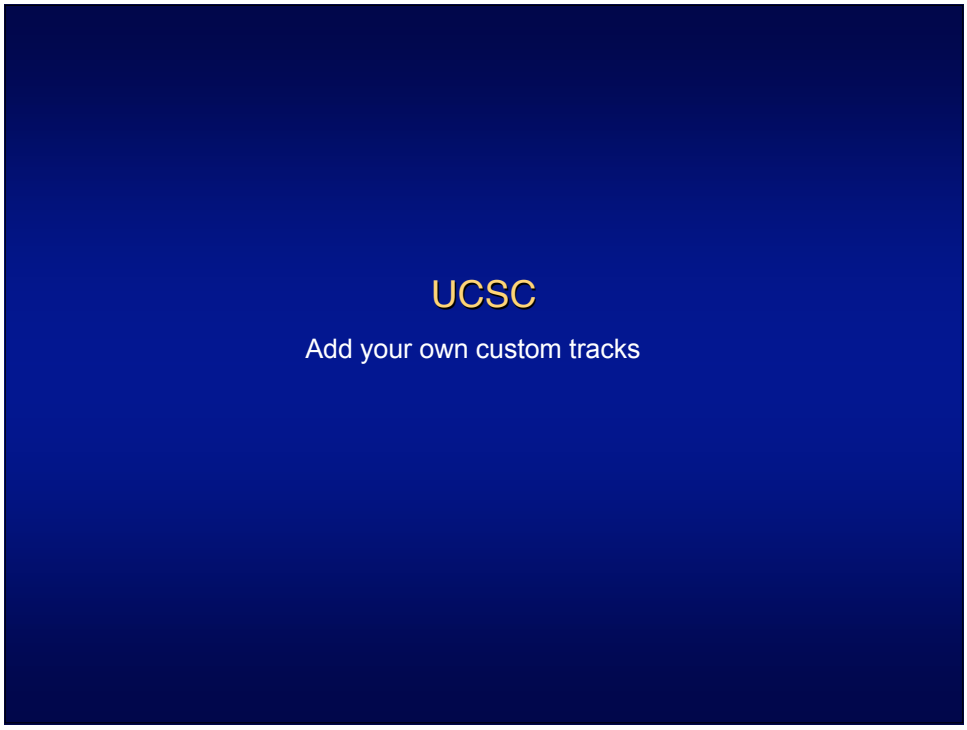
Find a chicken homolog of a human protein

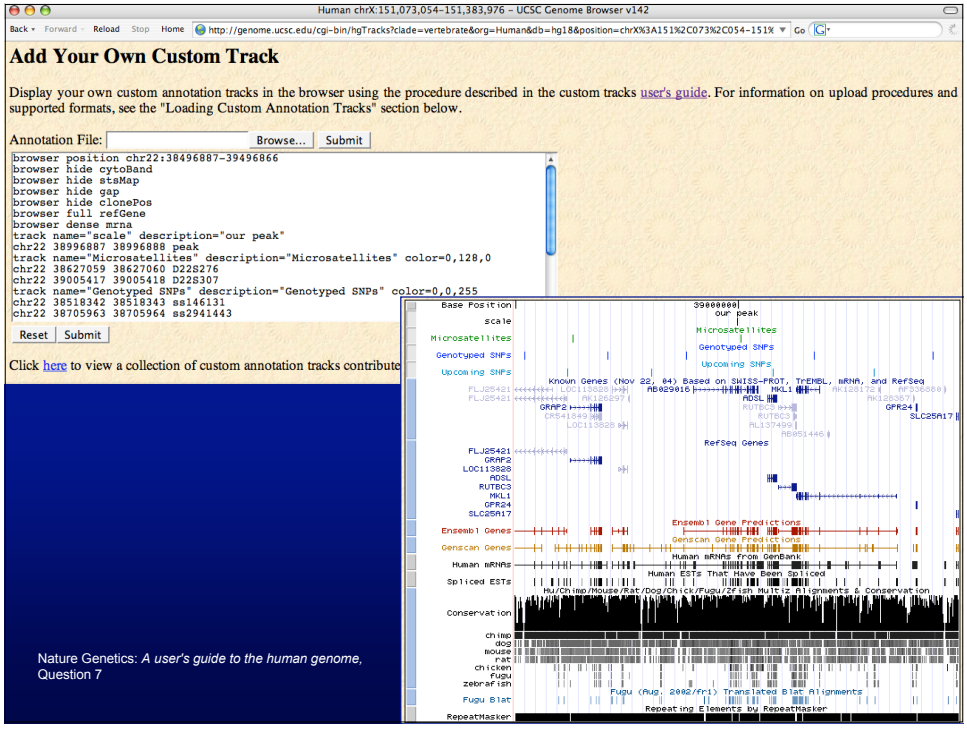



Chicken BLAT Search
http://genome.ucsc.edu/cgi-bin/hgBlat
UCSC BLAT search
Home Genomes Tables Gene Sorter FAQ Help
Chicken BLAT Search
BLAT Search Genome
Genome: Assembly: Query type: Sort output: Output type:
Chicken Feb. 2004 BLAT's guess query,score hyperlink
>gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domain 2 proprotein [Homo sapiens]
MWRVLFLLSGLGGLRMDSNFDSLPVQITVPEKIRSIIKEGIESQASYKVIIEGKPYTVNLMQKNFLPHNF
RVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQFENVSYGIEPLESSVGFEHVI
YQVKHKKADVSLYNEKDIESRDLSPKLQEVEPQQDFAKYIEMHVIVEKQLYNHMGSDTTVVAQKVFQLIG
LTNAIFVSFNITIILSSLELWIDENKIATTGEANELLHTFLRWKTSYLVLRPHDVAFLLVYREKSNYVGA
TFQGKMCDANYAGGVVLHPRTISLESLAVILAQLLSLSMGITYDDINKCQCSGAVCIMNPEAIHFSGVKI
FSNCSFEDFAHFISKQKSQCLHNQPRLDPFFKQQAVCGNAKLEAGEECDCGTEQDCALIGETCCDIATCR
FKAGSNCAEGPCCENCLFMSKERMCRPSFEECDLPEYCNGSSASCPENHYVQTGHPCGLNQWICIDGVCM
SGDKQCTDTFGKEVEPGPSECYSHLNGKIDVSGNCGISDSGYTQCEADNLQCGKLICKYVGKFLQIPRA
TIIYANISGHLCIAVEFASDHADSQKMWIKDGTSCGSNKVCRNQRCVSSSYLGYDCTTDKCNDRGVCNNK
KHCHCSASYLPPDCSVQSDLWPGGSIDSGNFPPVAIPARLPERRYIENIYHSKPMRWPFFLFIPFFIIFC
VLIAIMVKVNFQRKKWRTEDYSSDEQPESESEPKG
submit I'm feeling lucky clear
Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.
File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.
Upload sequence: Browse... submit file
Only DNA sequences of 25,000 or fewer bases and protein or translated sequence of 10000 or fewer letters will be processed. Up to 25 sequences can be submitted at the same time. The total limit for multiple sequence submissions is 50,000 bases or 25,000 letters.
For locating PCR primers, use In-Silico PCR for best results instead of BLAT.
About BLAT
BLAT on DNA is designed to quickly find sequences of 95% and greater similarity of length 40 bases or more. It may miss more divergent or shorter sequence alignments. It will find perfect sequence matches of 33 bases, and sometimes find them down to 20 bases. BLAT on proteins finds sequences of 80% and greater similarity of length 20 amino acids or more. In practice DNA BLAT works well on primates, and protein blat on land vertebrates.


Chicken BLAT Results
http://genome.ucsc.edu/cgi-bin/hgBlat
UCSC BLAT search
Home Genomes Tables Gene Sorter PCR FAQ Help
Chicken BLAT Results
BLAT Search Results
ACTIONS QUERY SCORE START END QSIZE IDENTITY CHRO STRAND START END SPAN
browser details NP_001455.3 44 539 600 735 71.6% Un ++ 635370 635555 186
browser details NP_001455.3 12 301 304 735 100.0% 1 ++ 67659709 67659720 12
browser details NP_001455.3 12 437 440 735 100.0% 1 ++ 67660117 67660128 12
browser details NP_001455.3 12 385 390 735 83.4% 1 ++ 67659961 67659978 18
UCSC Genome Browser on Chicken Feb. 2004 Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position/search chrUn:635,370-635,555 jump clear size 186 bp. configure
Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.
move start move end
default tracks hide all add custom tracks configure refresh
Chromosome Color Key:
Use drop down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
Mapping and Sequencing Tracks
Base Position Supercontigs Assembly Gap BAC End Pairs
GC Percent Quality Scores Isochores Contamination Short Match


Chicken BLAT Results
http://genome.ucsc.edu/cgi-bin/hgBlat
UCSC BLAT search
Home Genomes Tables Gene Sorter PCR FAQ Help
Chicken BLAT Results
BLAT Search Results
ACTIONS QUERY SCORE START END QSIZE IDENTITY CHRO STRAND START END SPAN
browser details NP_001455.3 44 539 600 735 71.6% Un ++ 635370 635555 186
browser details NP_001455.3 12 301 304 735 100.0% 1 ++ 67659709 67659720 12
browser details NP_001455.3 12 437 440 735 100.0% 1 ++ 67660117 67660128 12
browser details NP_0014
Alignment of NP_001455.3
NP_001455.3
Chicken.chrUn
block1
block2
block3
together
Alignment of NP_001455.3 and chrUn:635370-635555
Click on links in the frame to the left to navigate through the alignment. Matching bases are colored blue ar
sequence.
NP_001455.3
Chicken.chrUn :
Side by Side Alignment*


UCSC
Add your own custom tracks



UCSC Table Browser

- Download track in text format
- Retrieve DNA sequence covered by a track
- Calculate intersections between tracks and view in the Genome Browser. For example:
  - Show all RefSeq genes that contain only one exon
  - Show transcription factor binding sites that overlap (intersect) with a SNP

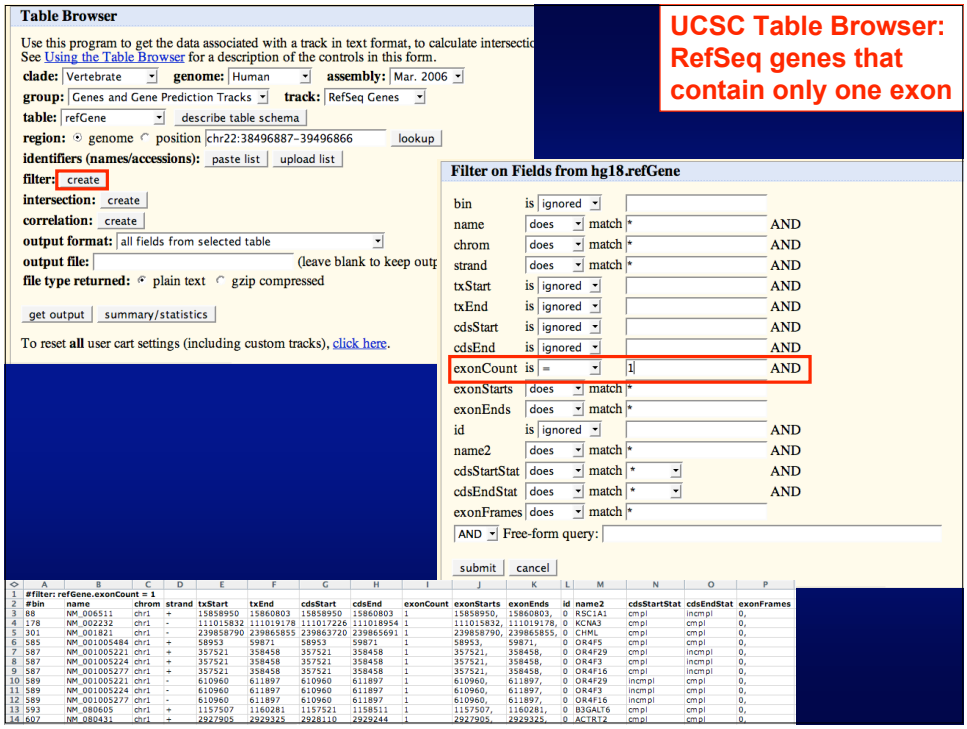
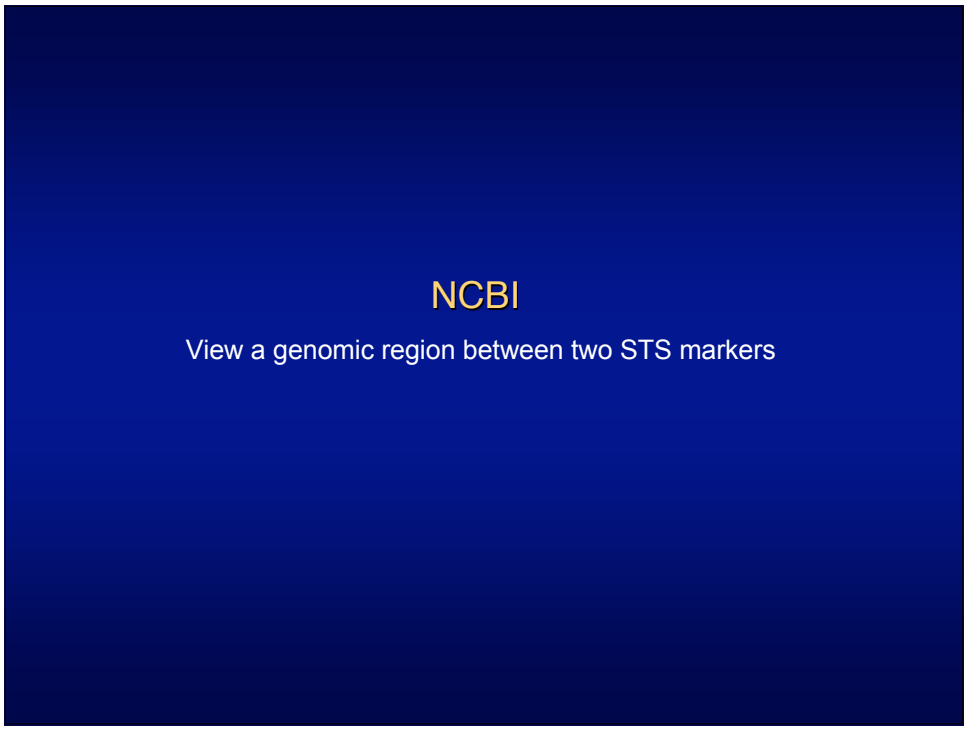
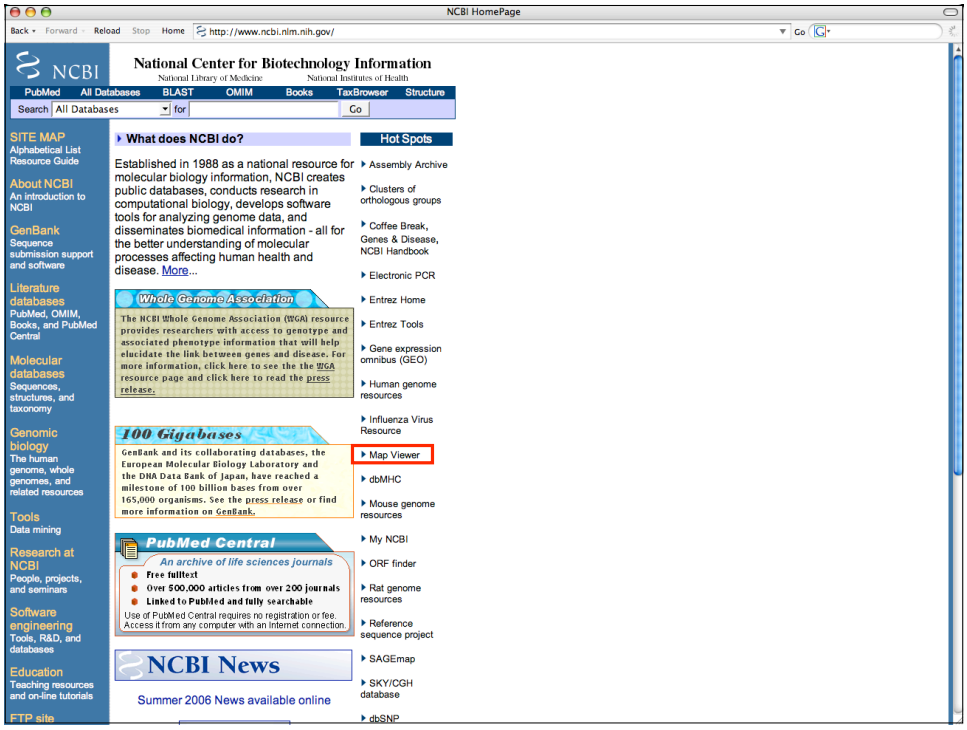
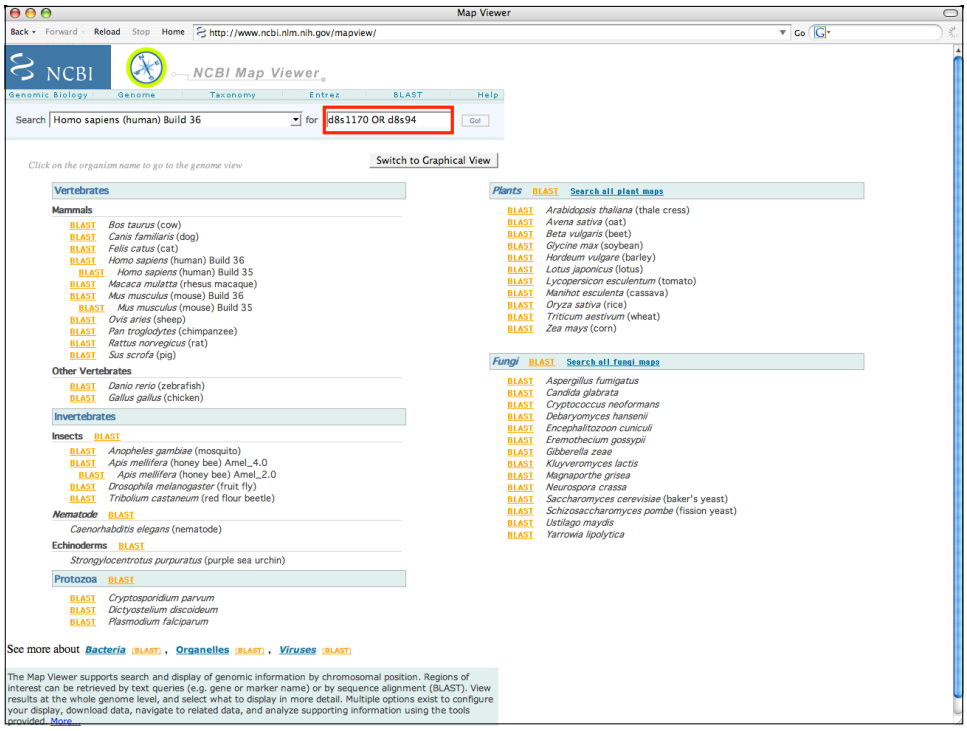
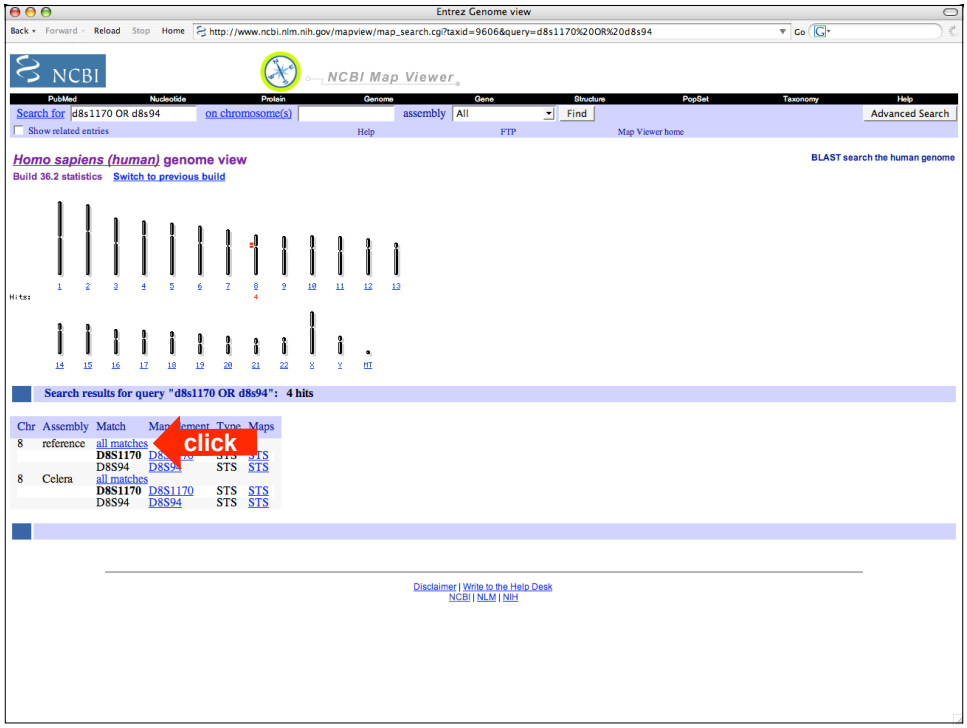


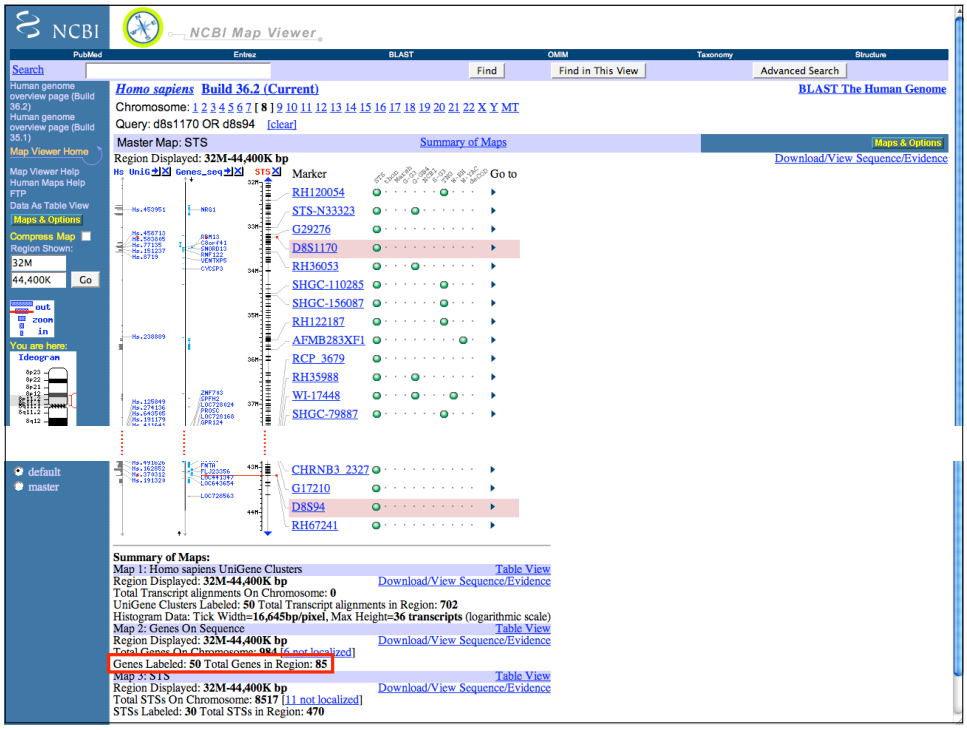
Table Browser
Use this program to get the data associated with a track in text format, to calculate intersectio
See Using the Table Browser for a description of the controls in this form.
clade: Vertebrate genome: Human assembly: Mar. 2006
group: Genes and Gene Prediction Tracks track: RefSeq Genes
table: refGene describe table schema
region: genome position chr22:38496887-39496866 lookup
identifiers (names/accessions): paste list upload list
filter: create
intersection: create
correlation: create
output format: all fields from selected table
output file: (leave blank to keep outp
file type returned: plain text gzip compressed
get output summary/statistics
To reset all user cart settings (including custom tracks), click here.
UCSC Table Browser: RefSeq genes that contain only one exon
Filter on Fields from hg18.refGene
bin is ignored
name does match * AND
chrom does match * AND
strand does match * AND
txStart is ignored AND
txEnd is ignored AND
cdsStart is ignored AND
cdsEnd is ignored AND
exonCount is = 1 AND
exonStarts does match *
exonEnds does match *
id is ignored AND
name2 does match * AND
cdsStartStat does match * AND
cdsEndStat does match * AND
exonFrames does match *
AND Free-form query:
submit cancel
#filter: refGene.exonCount = 1
#bin name chrom strand txStart txEnd cdsStart cdsEnd exonCount exonStarts exonEnds id name2 cdsStartStat cdsEndStat exonFrames
88 NM_006511 chr1 + 15858950 15860803 15858950 15860803 1 15858950, 15860803, 0 RSC1A1 cmpl incmpl 0,
178 NM_002232 chr1 - 111015832 111019178 111017226 111018954 1 111015832, 111019178, 0 KCNA3 cmpl cmpl 0,
301 NM_001821 chr1 - 239858790 239865855 239863720 239865691 1 239858790, 239865855, 0 CHML cmpl cmpl 0,
585 NM_001005484 chr1 + 58953 59871 58953 59871 1 58953, 59871, 0 OR4F5 cmpl cmpl 0,
587 NM_001005221 chr1 + 357521 358458 357521 358458 1 357521, 358458, 0 OR4F29 cmpl incmpl 0,
587 NM_001005224 chr1 + 357521 358458 357521 358458 1 357521, 358458, 0 OR4F3 cmpl incmpl 0,
587 NM_001005277 chr1 + 357521 358458 357521 358458 1 357521, 358458, 0 OR4F16 cmpl incmpl 0,
589 NM_001005221 chr1 - 610960 611897 610960 611897 1 610960, 611897, 0 OR4F29 incmpl cmpl 0,
589 NM_001005224 chr1 - 610960 611897 610960 611897 1 610960, 611897, 0 OR4F3 incmpl cmpl 0,
589 NM_001005277 chr1 - 610960 611897 610960 611897 1 610960, 611897, 0 OR4F16 incmpl cmpl 0,
593 NM_080605 chr1 + 1157507 1160281 1157521 1158511 1 1157507, 1160281, 0 B3GALT6 cmpl cmpl 0,
607 NM_080431 chr1 + 2927905 2929325 2928110 2929244 1 2927905, 2929325, 0 ACTRT2 cmpl cmpl 0,


NCBI
View a genomic region between two STS markers


NCBI HomePage
http://www.ncbi.nlm.nih.gov/
NCBI
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
PubMed All Databases BLAST OMIM Books TaxBrowser Structure
Search All Databases for Go
SITE MAP
Alphabetical List
Resource Guide
About NCBI
An introduction to NCBI
GenBank
Sequence submission support and software
Literature databases
PubMed, OMIM, Books, and PubMed Central
Molecular databases
Sequences, structures, and taxonomy
Genomic biology
The human genome, whole genomes, and related resources
Tools
Data mining
Research at NCBI
People, projects, and seminars
Software engineering
Tools, R&D, and databases
Education
Teaching resources and on-line tutorials
FTP site
What does NCBI do?
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...
Whole Genome Association
The NCBI Whole Genome Association (WGA) resource provides researchers with access to genotype and associated phenotype information that will help elucidate the link between genes and disease. For more information, click here to see the the WGA resource page and click here to read the press release.
100 Gigabases
GenBank and its collaborating databases, the European Molecular Biology Laboratory and the DNA Data Bank of Japan, have reached a milestone of 100 billion bases from over 165,000 organisms. See the press release or find more information on GenBank.
PubMed Central
An archive of life sciences journals
Free fulltext
Over 500,000 articles from over 200 journals
Linked to PubMed and fully searchable
Use of PubMed Central requires no registration or fee. Access it from any computer with an Internet connection.
NCBI News
Summer 2006 News available online
Hot Spots
Assembly Archive
Clusters of orthologous groups
Coffee Break, Genes & Disease, NCBI Handbook
Electronic PCR
Entrez Home
Entrez Tools
Gene expression omnibus (GEO)
Human genome resources
Influenza Virus Resource
Map Viewer
dbMHC
Mouse genome resources
My NCBI
ORF finder
Rat genome resources
Reference sequence project
SAGEmap
SKY/CGH database
dbSNP

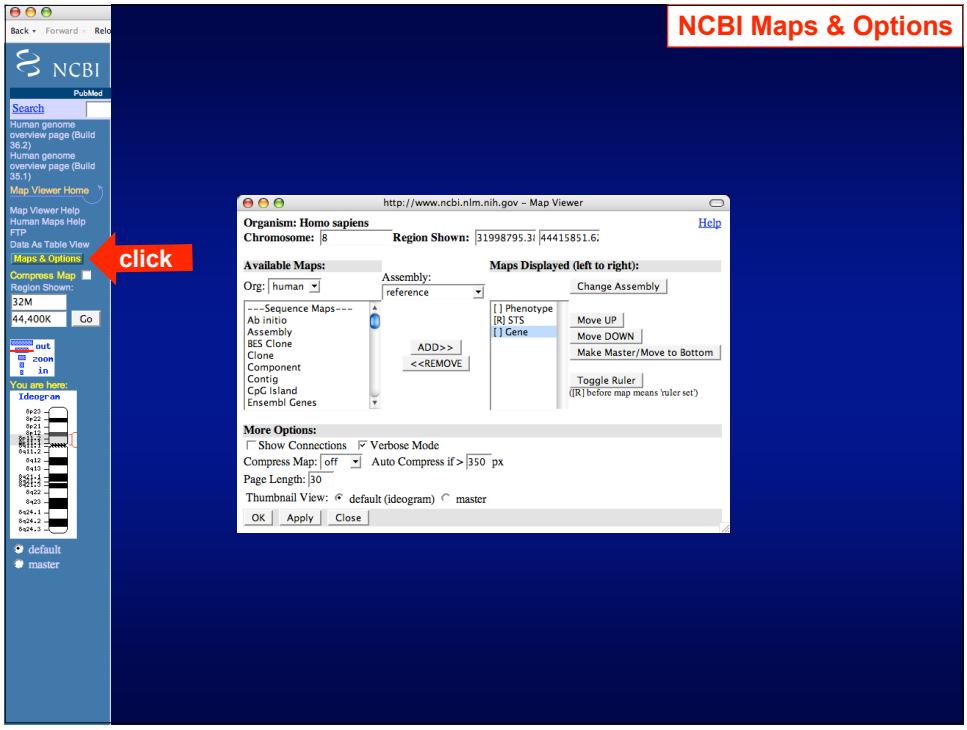

Map Viewer
http://www.ncbi.nlm.nih.gov/mapview/
NCBI
NCBI Map Viewer
Genomic Biology Genome Taxonomy Entrez BLAST Help
Search Homo sapiens (human) Build 36 for d8s1170 OR d8s94 Go!
Click on the organism name to go to the genome view
Switch to Graphical View
Vertebrates
Mammals
BLAST Bos taurus (cow)
BLAST Canis familiaris (dog)
BLAST Felis catus (cat)
BLAST Homo sapiens (human) Build 36
BLAST Homo sapiens (human) Build 35
BLAST Macaca mulatta (rhesus macaque)
BLAST Mus musculus (mouse) Build 36
BLAST Mus musculus (mouse) Build 35
BLAST Ovis aries (sheep)
BLAST Pan troglodytes (chimpanzee)
BLAST Rattus norvegicus (rat)
BLAST Sus scrofa (pig)
Other Vertebrates
BLAST Danio rerio (zebrafish)
BLAST Gallus gallus (chicken)
Invertebrates
Insects BLAST
BLAST Anopheles gambiae (mosquito)
BLAST Apis mellifera (honey bee) Amel_4.0
BLAST Apis mellifera (honey bee) Amel_2.0
BLAST Drosophila melanogaster (fruit fly)
BLAST Tribolium castaneum (red flour beetle)
Nematode BLAST
Caenorhabditis elegans (nematode)
Echinoderms BLAST
Strongylocentrotus purpuratus (purple sea urchin)
Protozoa BLAST
BLAST Cryptosporidium parvum
BLAST Dictyostelium discoideum
BLAST Plasmodium falciparum
Plants BLAST Search all plant maps
BLAST Arabidopsis thaliana (thale cress)
BLAST Avena sativa (oat)
BLAST Beta vulgaris (beet)
BLAST Glycine max (soybean)
BLAST Hordeum vulgare (barley)
BLAST Lotus japonicus (lotus)
BLAST Lycopersicon esculentum (tomato)
BLAST Manihot esculenta (cassava)
BLAST Oryza sativa (rice)
BLAST Triticum aestivum (wheat)
BLAST Zea mays (corn)
Fungi BLAST Search all fungi maps
BLAST Aspergillus fumigatus
BLAST Candida glabrata
BLAST Cryptococcus neoformans
BLAST Debaryomyces hansenii
BLAST Encephalitozoon cuniculi
BLAST Eremothecium gossypii
BLAST Gibberella zeae
BLAST Kluyveromyces lactis
BLAST Magnaporthe grisea
BLAST Neurospora crassa
BLAST Saccharomyces cerevisiae (baker's yeast)
BLAST Schizosaccharomyces pombe (fission yeast)
BLAST Ustilago maydis
BLAST Yarrowia lipolytica
See more about Bacteria (BLAST), Organelles (BLAST), Viruses (BLAST)
The Map Viewer supports search and display of genomic information by chromosomal position. Regions of interest can be retrieved by text queries (e.g. gene or marker name) or by sequence alignment (BLAST). View results at the whole genome level, and select what to display in more detail. Multiple options exist to configure your display, download data, navigate to related data, and analyze supporting information using the tools provided. More...

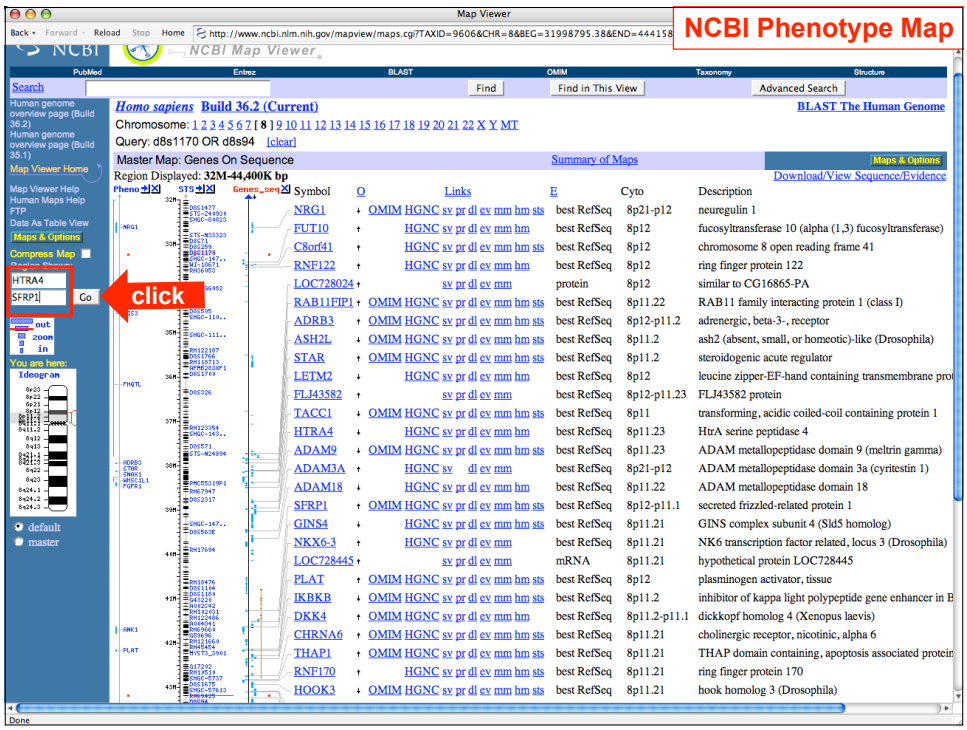

Entrez Genome view
NCBI
NCBI Map Viewer
Search for d8s1170 OR d8s94
Homo sapiens (human) genome view
Search results for query "d8s1170 OR d8s94": 4 hits
click

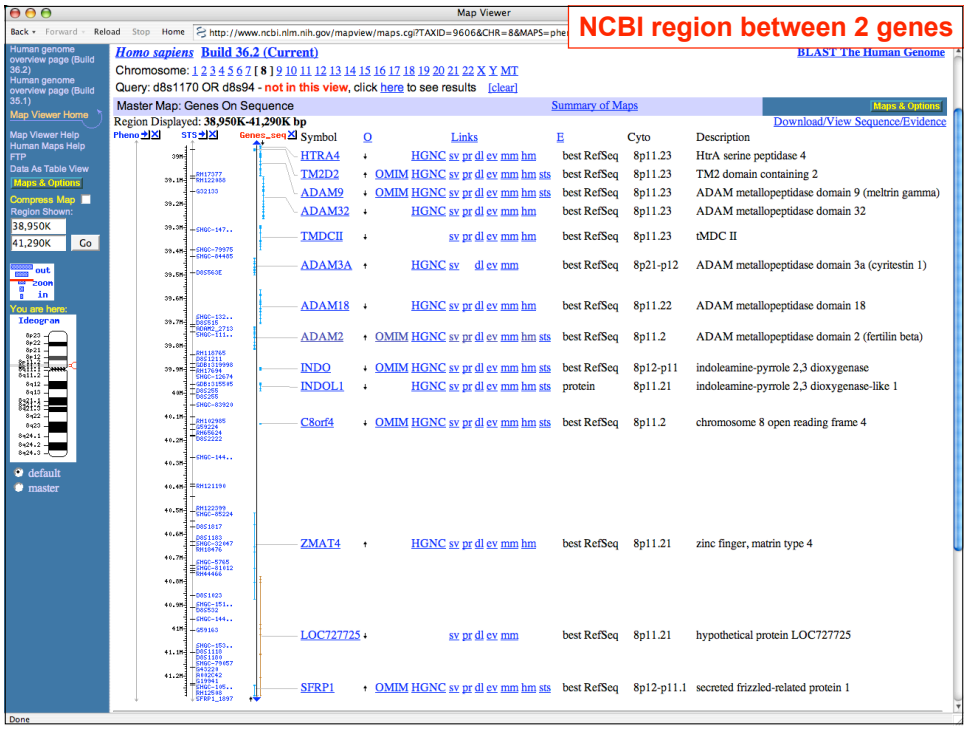

NCBI
NCBI Map Viewer
Homo sapiens Build 36.2 (Current)
Query: d8s1170 OR d8s94
Master Map: STS
Region Displayed: 32M-44,400K bp
Summary of Maps:

NCBI

Change the maps displayed on the Map Viewer

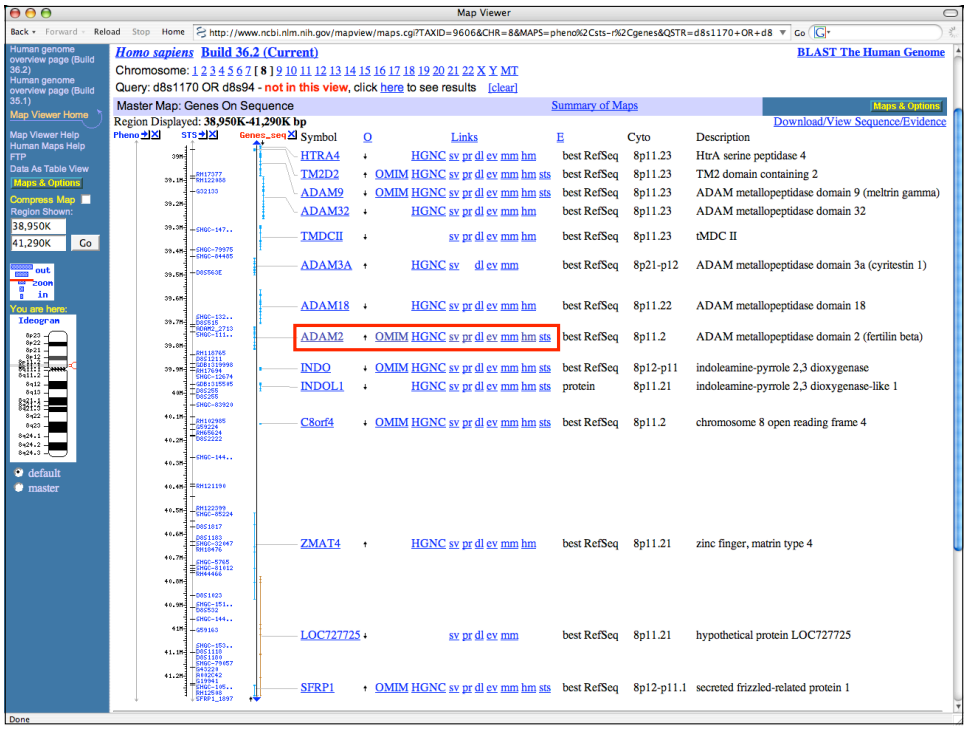


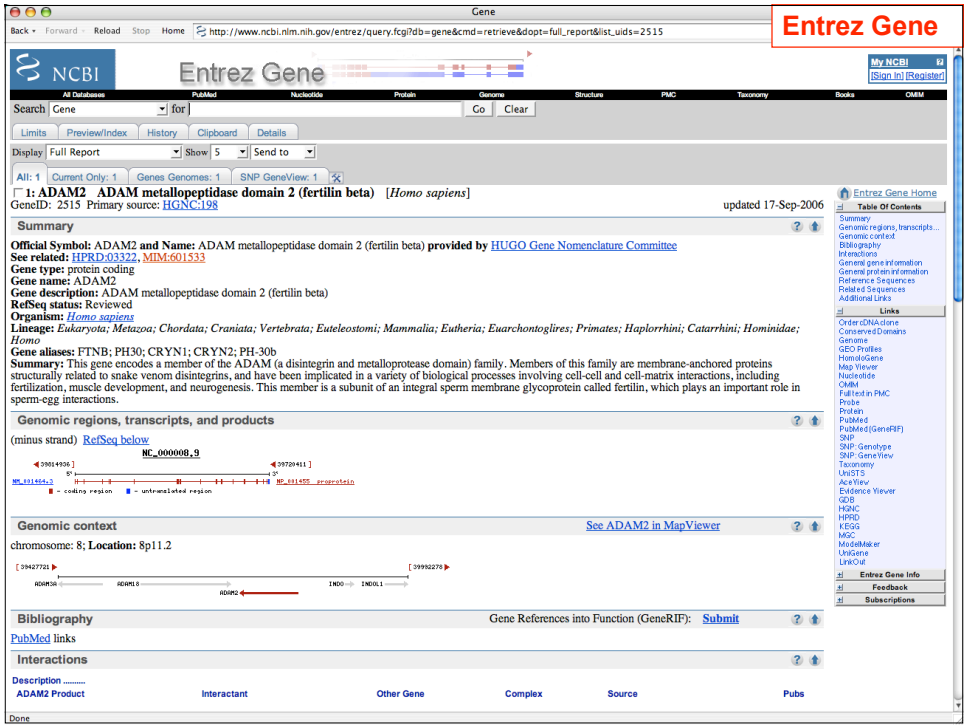

NCBI Phenotype Map
Map Viewer
NCBI Map Viewer
Homo sapiens Build 36.2 (Current)
BLAST The Human Genome
Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: d8s1170 OR d8s94 [clear]
Master Map: Genes On Sequence
Summary of Maps
Maps & Options
Region Displayed: 32M-44,400K bp
Download/View Sequence/Evidence
HTRA4
SFRP1
Go
click
Symbol O Links E Cyto Description
NRG1 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p21-p12 neuregulin 1
FUT10 HGNC sv pr dl ev mm hm best RefSeq 8p12 fucosyltransferase 10 (alpha (1,3) fucosyltransferase)
C8orf41 HGNC sv pr dl ev mm hm sts best RefSeq 8p12 chromosome 8 open reading frame 41
RNF122 HGNC sv pr dl ev mm hm best RefSeq 8p12 ring finger protein 122
LOC728024 sv pr dl ev mm protein 8p12 similar to CG16865-PA
RAB11FIP1 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.22 RAB11 family interacting protein 1 (class I)
ADRB3 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p12-p11.2 adrenergic, beta-3-, receptor
ASH2L OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.2 ash2 (absent, small, or homeotic)-like (Drosophila)
STAR OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.2 steroidogenic acute regulator
LETM2 HGNC sv pr dl ev mm hm best RefSeq 8p12 leucine zipper-EF-hand containing transmembrane pro
FLJ43582 sv pr dl ev mm best RefSeq 8p12-p11.23 FLJ43582 protein
TACC1 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11 transforming, acidic coiled-coil containing protein 1
HTRA4 HGNC sv pr dl ev mm hm best RefSeq 8p11.23 HtrA serine peptidase 4
ADAM9 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.23 ADAM metallopeptidase domain 9 (meltrin gamma)
ADAM3A HGNC sv dl ev mm best RefSeq 8p21-p12 ADAM metallopeptidase domain 3a (cyritestin 1)
ADAM18 HGNC sv pr dl ev mm hm best RefSeq 8p11.22 ADAM metallopeptidase domain 18
SFRP1 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p12-p11.1 secreted frizzled-related protein 1
GINS4 HGNC sv pr dl ev mm hm sts best RefSeq 8p11.21 GINS complex subunit 4 (Sld5 homolog)
NKX6-3 HGNC sv pr dl ev mm best RefSeq 8p11.21 NK6 transcription factor related, locus 3 (Drosophila)
LOC728445 sv pr dl ev mm mRNA 8p11.21 hypothetical protein LOC728445
PLAT OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p12 plasminogen activator, tissue
IKBKB OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.2 inhibitor of kappa light polypeptide gene enhancer in B
DKK4 OMIM HGNC sv pr dl ev mm hm best RefSeq 8p11.2-p11.1 dickkopf homolog 4 (Xenopus laevis)
CHRNA6 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.21 cholinergic receptor, nicotinic, alpha 6
THAP1 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.21 THAP domain containing, apoptosis associated protein
RNF170 HGNC sv pr dl ev mm hm sts best RefSeq 8p11.21 ring finger protein 170
HOOK3 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.21 hook homolog 3 (Drosophila)

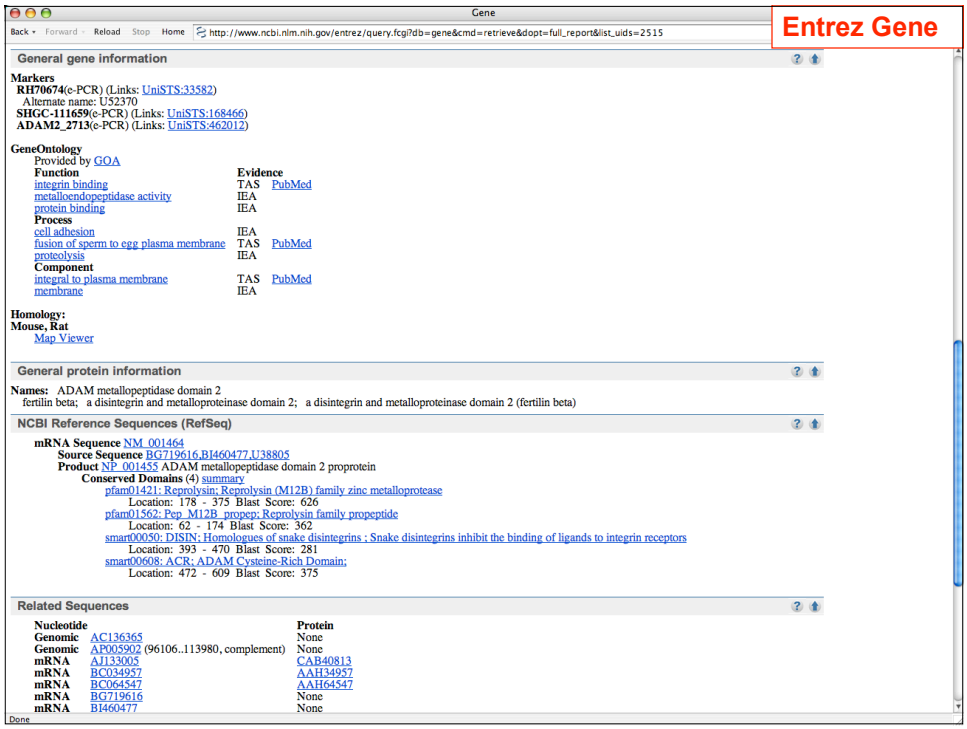

NCBI region between 2 genes
Map Viewer
Homo sapiens Build 36.2 (Current)
BLAST The Human Genome
Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: d8s1170 OR d8s94 - not in this view, click here to see results [clear]
Master Map: Genes On Sequence
Summary of Maps
Maps & Options
Region Displayed: 38,950K-41,290K bp
Download/View Sequence/Evidence
38,950K
41,290K
Go
Symbol O Links E Cyto Description
HTRA4 HGNC sv pr dl ev mm hm best RefSeq 8p11.23 HtrA serine peptidase 4
TM2D2 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.23 TM2 domain containing 2
ADAM9 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.23 ADAM metallopeptidase domain 9 (meltrin gamma)
ADAM32 HGNC sv pr dl ev mm hm best RefSeq 8p11.23 ADAM metallopeptidase domain 32
TMDCII sv pr dl ev mm hm best RefSeq 8p11.23 tMDC II
ADAM3A HGNC sv dl ev mm best RefSeq 8p21-p12 ADAM metallopeptidase domain 3a (cyritestin 1)
ADAM18 HGNC sv pr dl ev mm hm best RefSeq 8p11.22 ADAM metallopeptidase domain 18
ADAM2 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.2 ADAM metallopeptidase domain 2 (fertilin beta)
INDO OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p12-p11 indoleamine-pyrrole 2,3 dioxygenase
INDOL1 HGNC sv pr dl ev mm hm sts protein 8p11.21 indoleamine-pyrrole 2,3 dioxygenase-like 1
C8orf4 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p11.2 chromosome 8 open reading frame 4
ZMAT4 HGNC sv pr dl ev mm hm best RefSeq 8p11.21 zinc finger, matrin type 4
LOC727725 sv pr dl ev mm best RefSeq 8p11.21 hypothetical protein LOC727725
SFRP1 OMIM HGNC sv pr dl ev mm hm sts best RefSeq 8p12-p11.1 secreted frizzled-related protein 1

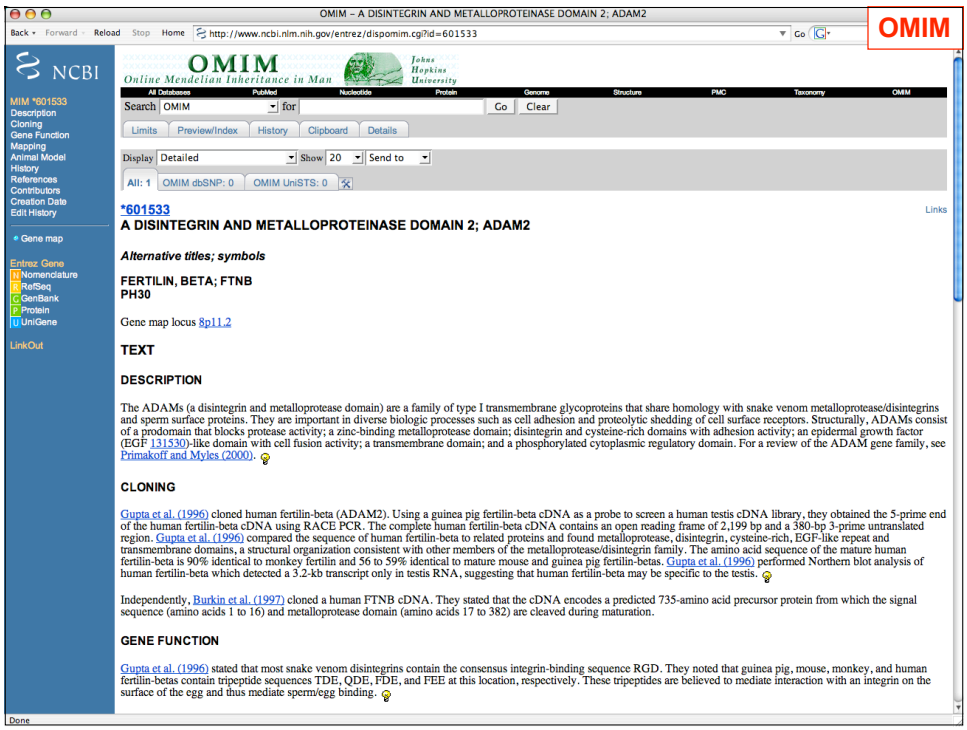
NCBI

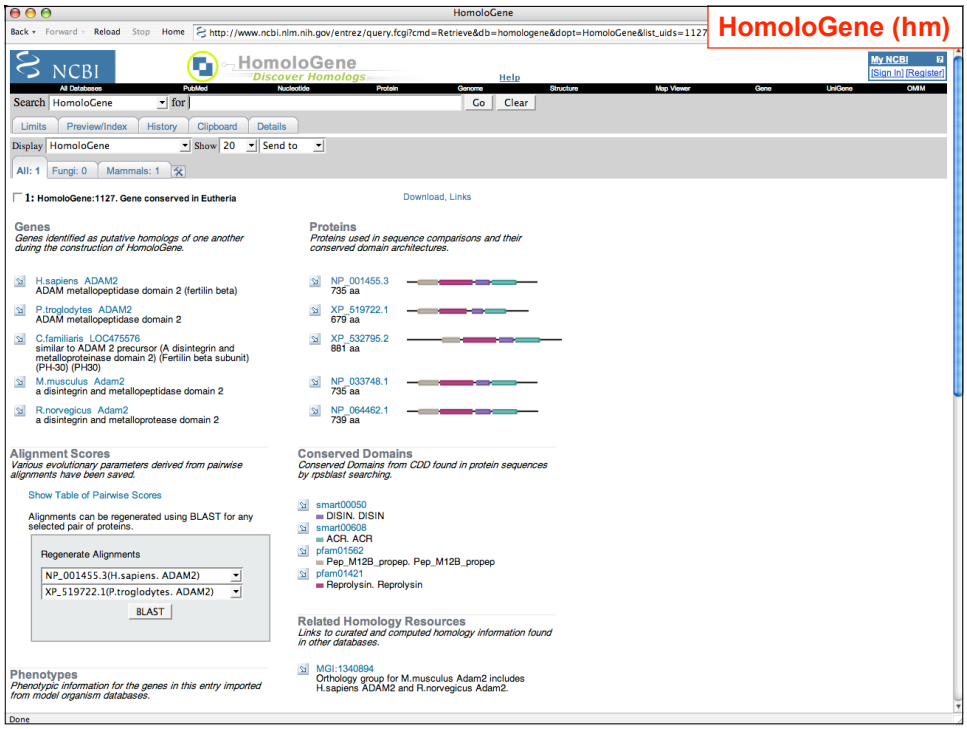
View additional information about a gene




Entrez Gene
Gene
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=full_report&list_uids=2515
NCBI
Entrez Gene
1: ADAM2 ADAM metallopeptidase domain 2 (fertilin beta) [Homo sapiens]
GeneID: 2515 Primary source: HGNC:198
updated 17-Sep-2006
Summary
Official Symbol: ADAM2 and Name: ADAM metallopeptidase domain 2 (fertilin beta) provided by HUGO Gene Nomenclature Committee
See related: HPRD:03322, MIM:601533
Gene type: protein coding
Gene name: ADAM2
Gene description: ADAM metallopeptidase domain 2 (fertilin beta)
RefSeq status: Reviewed
Organism: Homo sapiens
Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Gene aliases: FTNB; PH30; CRYN1; CRYN2; PH-30b
Summary: This gene encodes a member of the ADAM (a disintegrin and metalloprotease domain) family. Members of this family are membrane-anchored proteins structurally related to snake venom disintegrins, and have been implicated in a variety of biological processes involving cell-cell and cell-matrix interactions, including fertilization, muscle development, and neurogenesis. This member is a subunit of an integral sperm membrane glycoprotein called fertilin, which plays an important role in sperm-egg interactions.
Genomic regions, transcripts, and products
Genomic context
See ADAM2 in MapViewer
chromosome: 8; Location: 8p11.2
Bibliography
Gene References into Function (GeneRIF): Submit
PubMed links
Interactions
Entrez Gene Home
Table Of Contents
Links
Entrez Gene Info
Feedback
Subscriptions

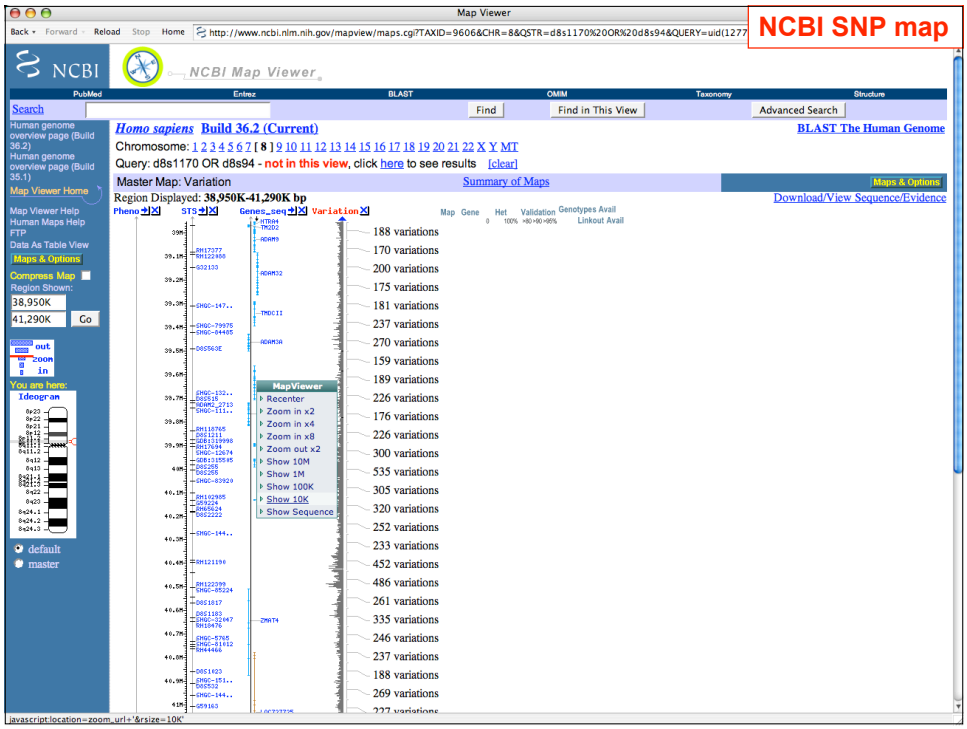

Entrez Gene
Gene
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=full_report&list_uids=2515
General gene information
Markers
RH70674(e-PCR) (Links: UniSTS:33582)
Alternate name: U52370
SHGC-111659(e-PCR) (Links: UniSTS:168466)
ADAM2_2713(e-PCR) (Links: UniSTS:462012)
GeneOntology
Provided by GOA
Function
integrin binding
metalloendopeptidase activity
protein binding
Process
cell adhesion
fusion of sperm to egg plasma membrane
proteolysis
Component
integral to plasma membrane
membrane
Evidence
TAS PubMed
IEA
IEA
IEA
TAS PubMed
IEA
TAS PubMed
IEA
Homology:
Mouse, Rat
Map Viewer
General protein information
Names: ADAM metallopeptidase domain 2
fertilin beta; a disintegrin and metalloproteinase domain 2; a disintegrin and metalloproteinase domain 2 (fertilin beta)
NCBI Reference Sequences (RefSeq)
mRNA Sequence NM_001464
Source Sequence BG719616,BI460477,U38805
Product NP_001455 ADAM metallopeptidase domain 2 proprotein
Conserved Domains (4) summary
pfam01421: Reprolysin; Reprolysin (M12B) family zinc metalloprotease
Location: 178 - 375 Blast Score: 626
pfam01562: Pep_M12B_propep; Reprolysin family propeptide
Location: 62 - 174 Blast Score: 362
smart00050: DISIN; Homologues of snake disintegrins ; Snake disintegrins inhibit the binding of ligands to integrin receptors
Location: 393 - 470 Blast Score: 281
smart00608: ACR; ADAM Cysteine-Rich Domain;
Location: 472 - 609 Blast Score: 375
Related Sequences
Nucleotide
Genomic AC136365
Genomic AP005902 (96106..113980, complement)
mRNA AJ133005
mRNA BC034957
mRNA BC064547
mRNA BG719616
mRNA BI460477
Protein
None
None
CAB40813
AAH34957
AAH64547
None
None

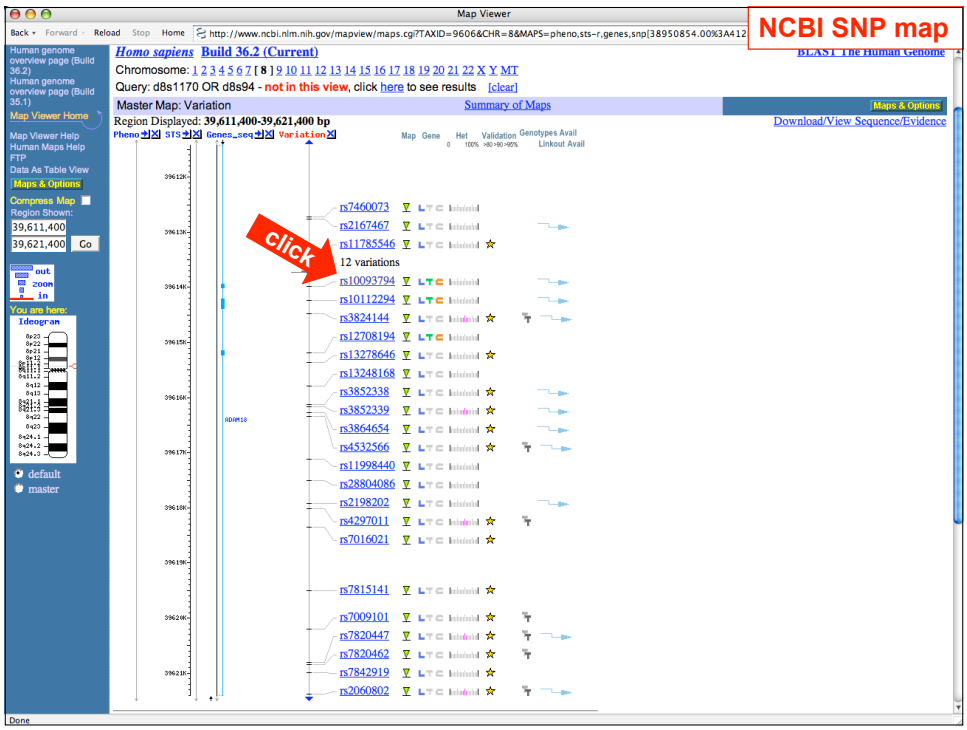

OMIM
OMIM – A DISINTEGRIN AND METALLOPROTEINASE DOMAIN 2; ADAM2
http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=601533
NCBI
OMIM
Online Mendelian Inheritance in Man
Johns Hopkins University
*601533
A DISINTEGRIN AND METALLOPROTEINASE DOMAIN 2; ADAM2
Alternative titles; symbols
FERTILIN, BETA; FTNB
PH30
Gene map locus 8p11.2
TEXT
DESCRIPTION
CLONING
GENE FUNCTION

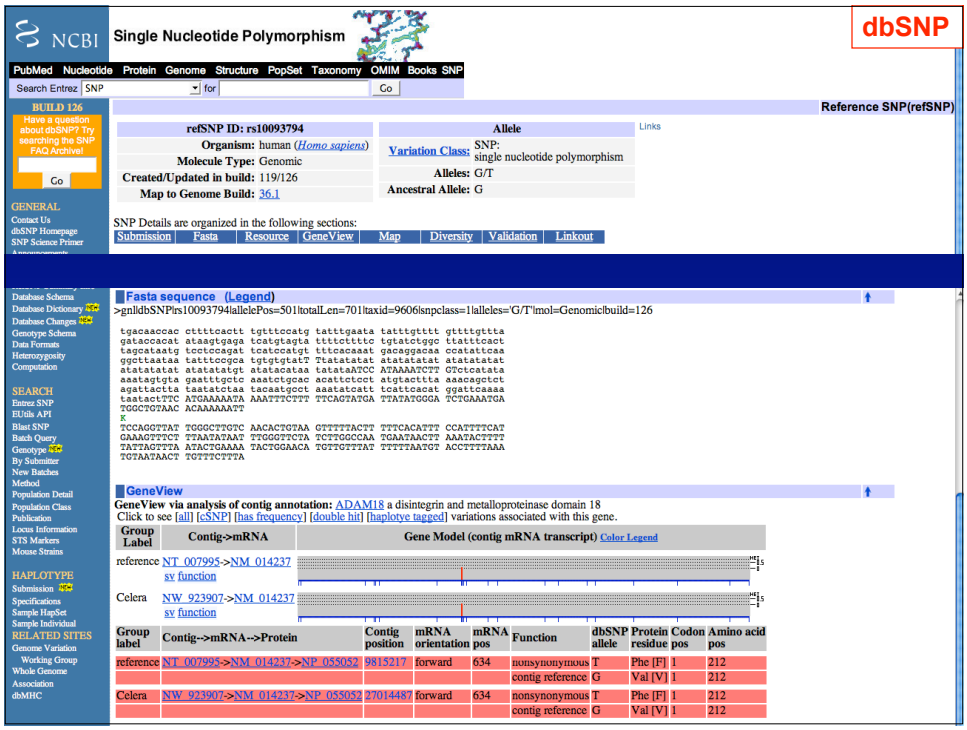

HomoloGene (hm)
HomoloGene
Discover Homologs
1: HomoloGene:1127. Gene conserved in Eutheria
Genes
Genes identified as putative homologs of one another during the construction of HomoloGene.
Proteins
Proteins used in sequence comparisons and their conserved domain architectures.
Alignment Scores
Various evolutionary parameters derived from pairwise alignments have been saved.
Conserved Domains
Conserved Domains from CDD found in protein sequences by rpsblast searching.
Related Homology Resources
Links to curated and computed homology information found in other databases.
Phenotypes
Phenotypic information for the genes in this entry imported from model organism databases.

NCBI

Zoom in to view finer detail

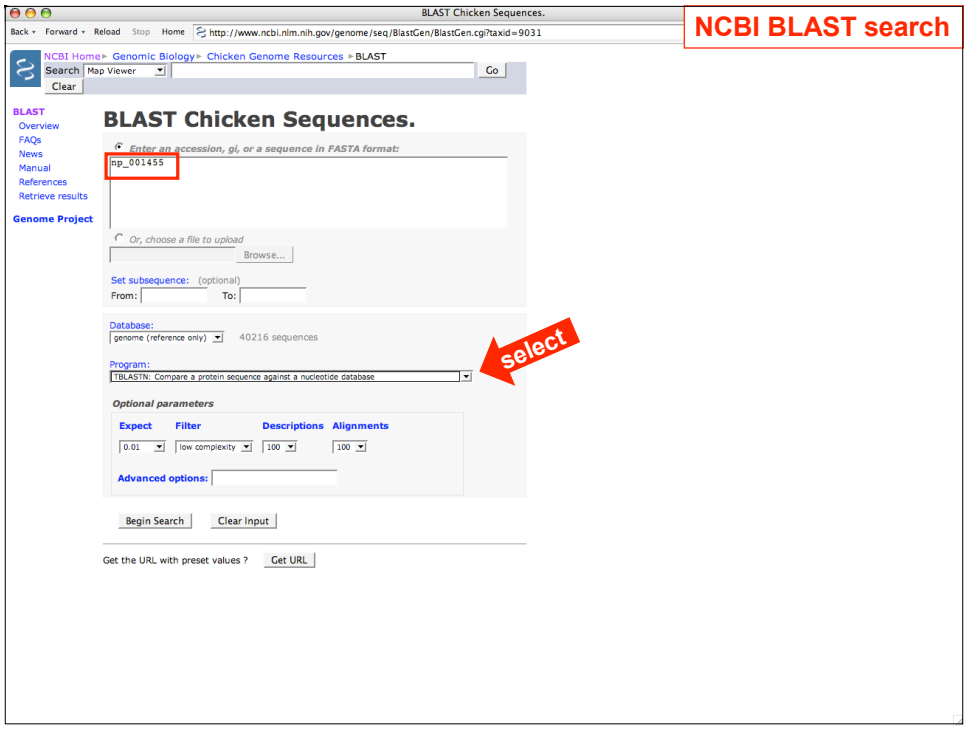


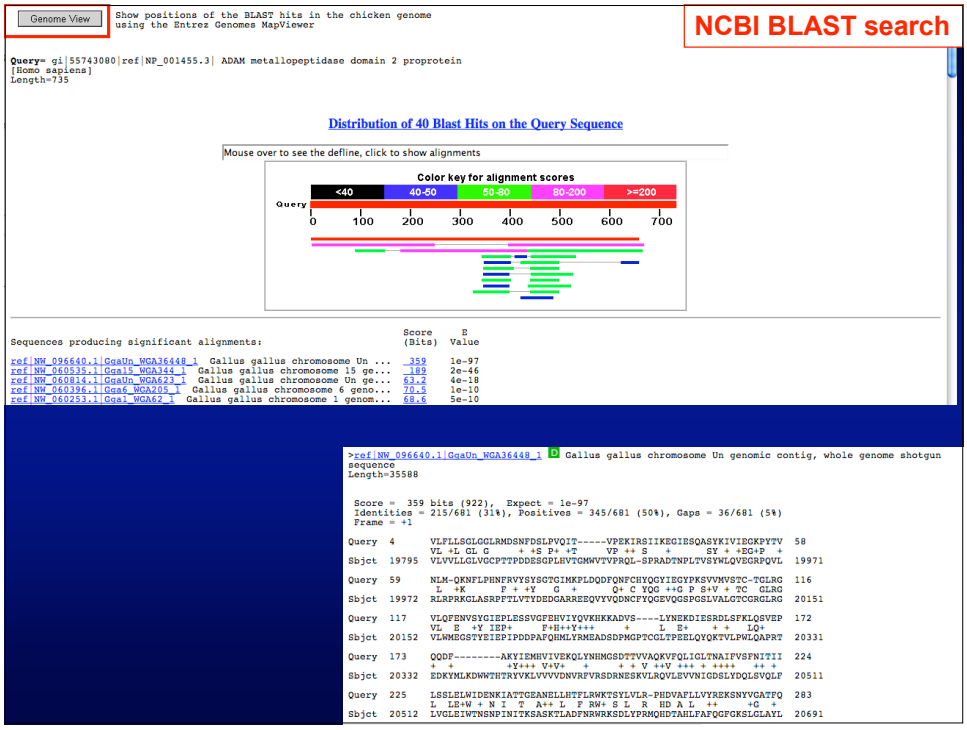

NCBI SNP map
Map Viewer
Homo sapiens Build 36.2 (Current)
Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: d8s1170 OR d8s94 - not in this view, click here to see results [clear]
Master Map: Variation
Summary of Maps
Maps & Options
Region Displayed: 39,611,400-39,621,400 bp
Download/View Sequence/Evidence
click
12 variations
rs7460073
rs2167467
rs11785546
rs10093794
rs10112294
rs3824144
rs12708194
rs13278646
rs13248168
rs3852338
rs3852339
rs3864654
rs4532566
rs11998440
rs28804086
rs2198202
rs4297011
rs7016021
rs7815141
rs7009101
rs7820447
rs7820462
rs7842919
rs2060802

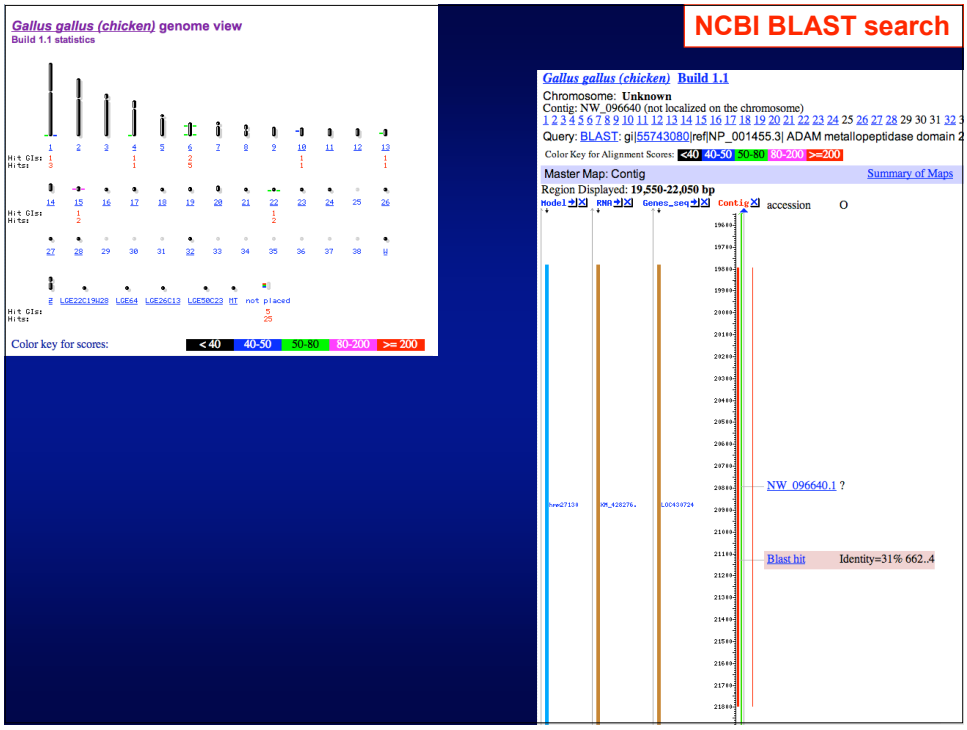

dbSNP
Single Nucleotide Polymorphism
Reference SNP(refSNP)
refSNP ID: rs10093794
Organism: human (Homo sapiens)
Molecule Type: Genomic
Created/Updated in build: 119/126
Map to Genome Build: 36.1
Allele
Variation Class: SNP: single nucleotide polymorphism
Alleles: G/T
Ancestral Allele: G
SNP Details are organized in the following sections:
Submission Fasta Resource GeneView Map Diversity Validation Linkout
Fasta sequence (Legend)
GeneView
GeneView via analysis of contig annotation: ADAM18 a disintegrin and metalloproteinase domain 18
reference NT_007995->NM_014237
Celera NW_923907->NM_014237
Group label
Contig->mRNA-->Protein
Contig position
mRNA orientation
mRNA pos
Function
dbSNP allele
Protein residue
Codon pos
Amino acid pos
reference NT_007995->NM_014237->NP_055052 9815217 forward 634 nonsynonymous T Phe [F] 1 212
contig reference G Val [V] 1 212
Celera NW_923907->NM_014237->NP_055052 27014487 forward 634 nonsynonymous T Phe [F] 1 212
contig reference G Val [V] 1 212

NCBI

Find a chicken homolog of a human protein

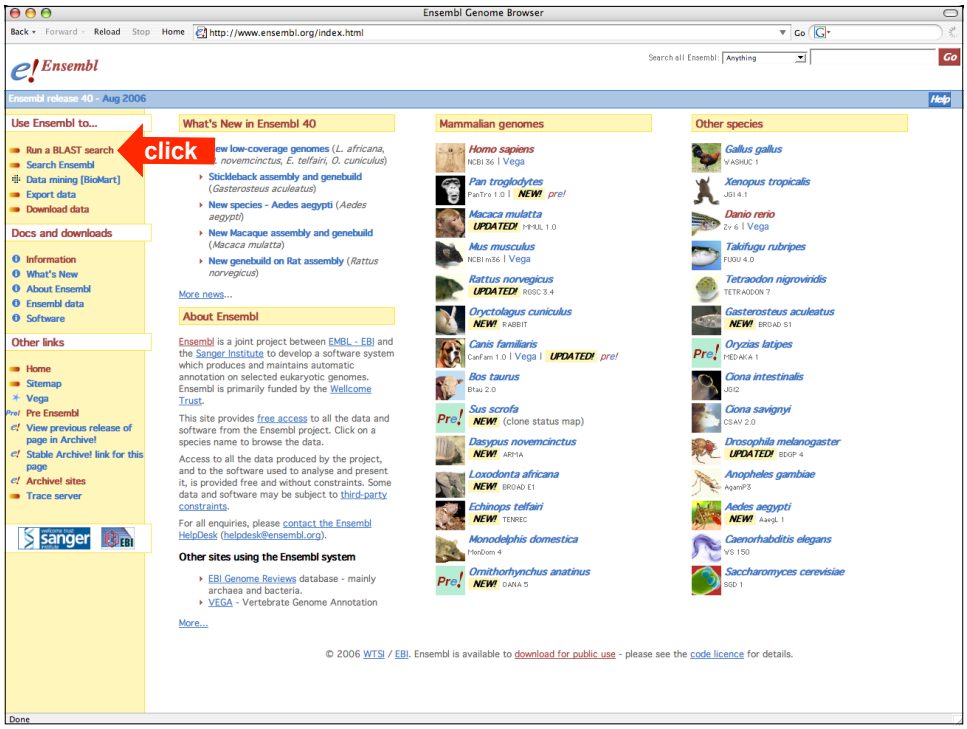


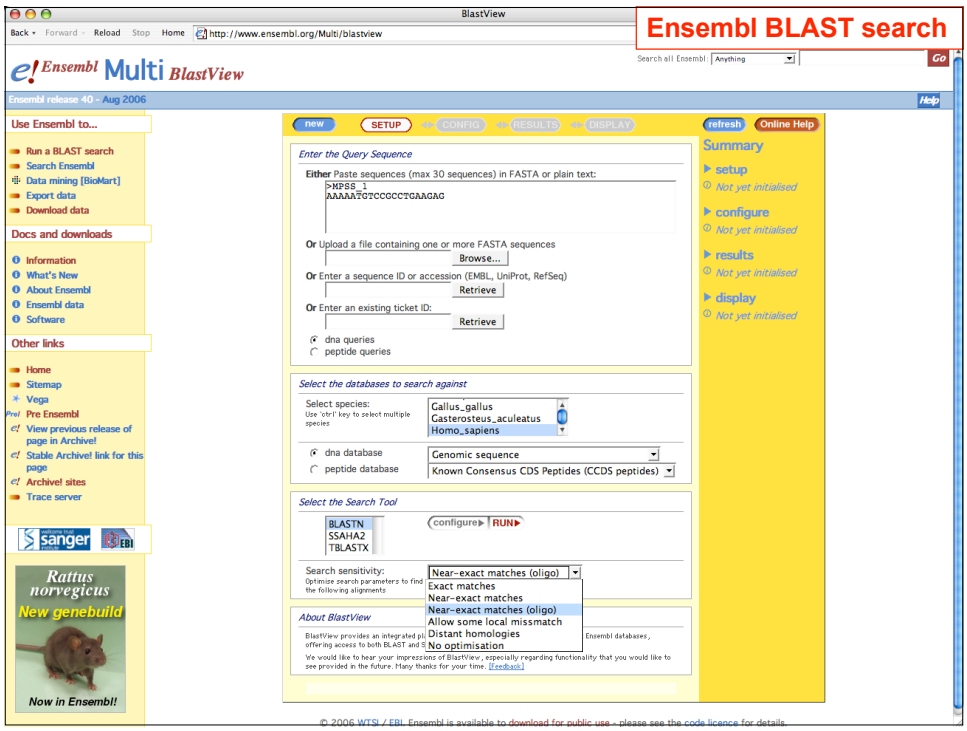

Genome View
Show positions of the BLAST hits in the chicken genome using the Entrez Genomes MapViewer
NCBI BLAST search
Query= gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domain 2 proprotein [Homo sapiens]
Length=735
Distribution of 40 Blast Hits on the Query Sequence
Mouse over to see the defline, click to show alignments
Color key for alignment scores
<40
40-50
50-80
80-200
>=200
Query
0
100
200
300
400
500
600
700
Sequences producing significant alignments:
Score (Bits)
E Value
ref|NW_096640.1|GgaUn_WGA36448_1 Gallus gallus chromosome Un ... 359 1e-97
ref|NW_060535.1|Gga15_WGA344_1 Gallus gallus chromosome 15 ge... 189 2e-46
ref|NW_060814.1|GgaUn_WGA623_1 Gallus gallus chromosome Un ge... 63.2 4e-18
ref|NW_060396.1|Gga6_WGA205_1 Gallus gallus chromosome 6 geno... 70.5 1e-10
ref|NW_060253.1|Gga1_WGA62_1 Gallus gallus chromosome 1 genom... 68.6 5e-10
>ref|NW_096640.1|GgaUn_WGA36448_1 Gallus gallus chromosome Un genomic contig, whole genome shotgun sequence
Length=35588
Score = 359 bits (922), Expect = 1e-97
Identities = 215/681 (31%), Positives = 345/681 (50%), Gaps = 36/681 (5%)
Frame = +1

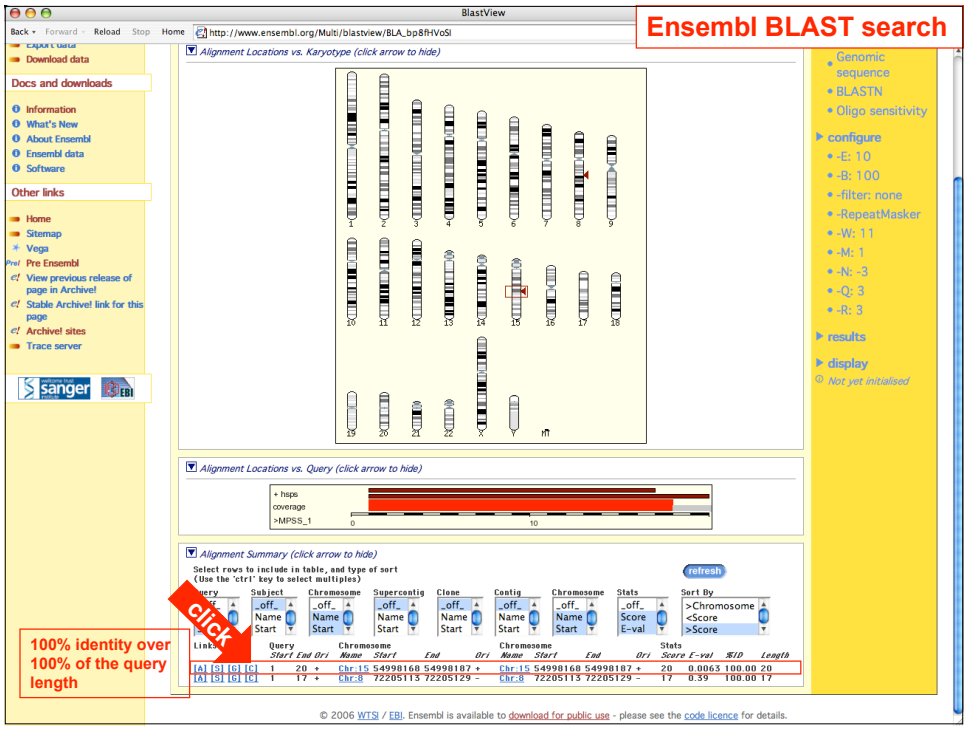

Gallus gallus (chicken) genome view
Build 1.1 statistics
Color key for scores:
< 40
40-50
50-80
80-200
>= 200
NCBI BLAST search
Gallus gallus (chicken) Build 1.1
Chromosome: Unknown
Contig: NW_096640 (not localized on the chromosome)
Query: BLAST: gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domain 2
Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200
Master Map: Contig
Summary of Maps
Region Displayed: 19,550-22,050 bp
Model
RNA
Genes_seq
Contig
accession
NW_096640.1 ?
Blast hit
Identity=31% 662..4

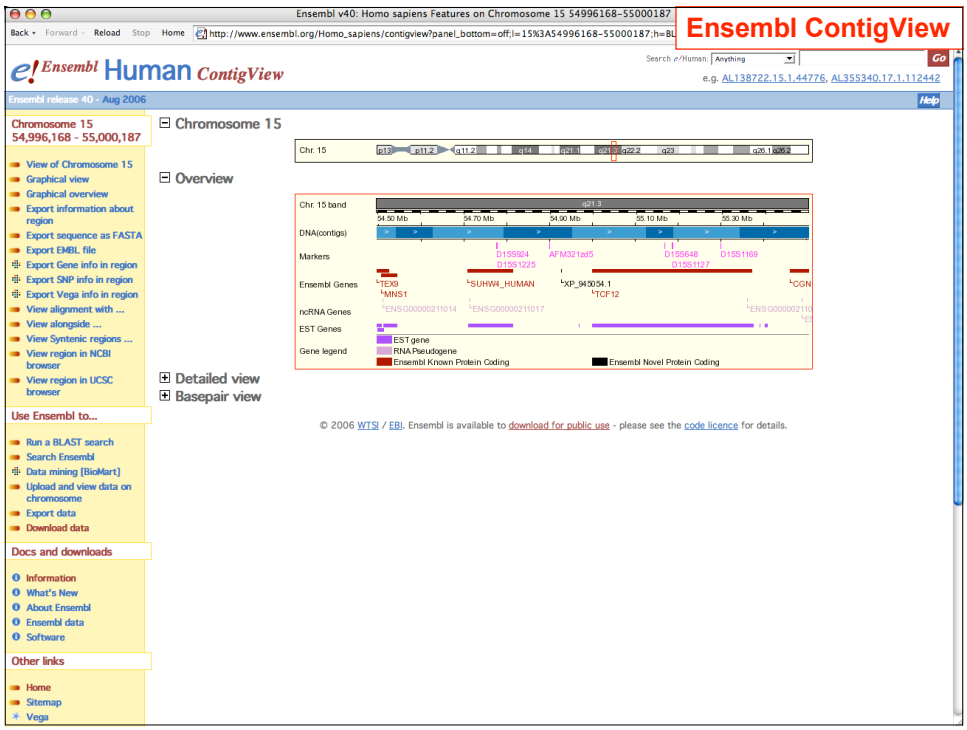
Ensembl

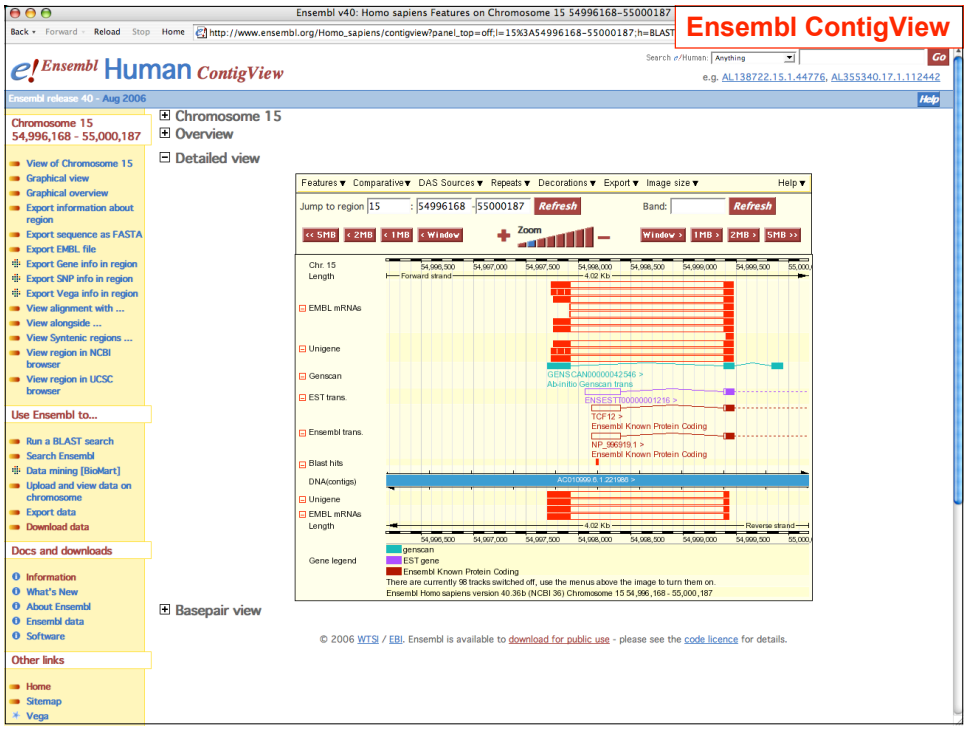
Identify genes that overlap with an oligo tag

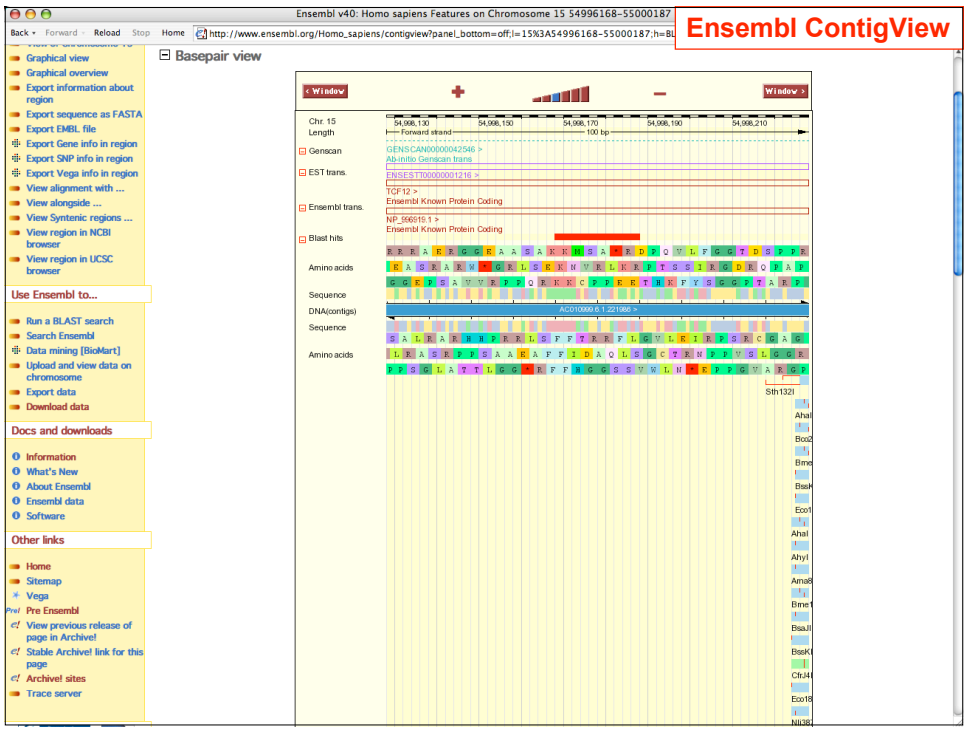


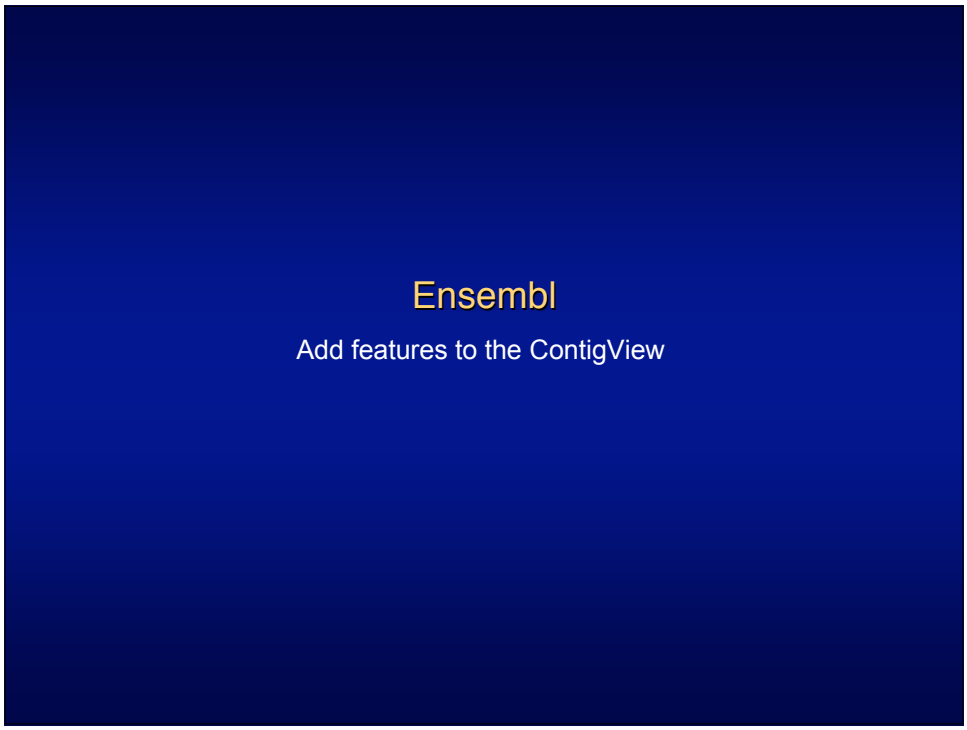

Ensembl BLAST search
BlastView
http://www.ensembl.org/Multi/blastview
Ensembl Multi BlastView
Ensembl release 40 - Aug 2006
Use Ensembl to...
Run a BLAST search
Search Ensembl
Data mining [BioMart]
Export data
Download data
Docs and downloads
Information
What's New
About Ensembl
Ensembl data
Software
Other links
Home
Sitemap
Vega
Pre Ensembl
View previous release of page in Archive!
Stable Archive! link for this page
Archive! sites
Trace server
Rattus norvegicus
New genebuild
Now in Ensembl!
new
SETUP
CONFIG
RESULTS
DISPLAY
refresh
Online Help
Enter the Query Sequence
Either Paste sequences (max 30 sequences) in FASTA or plain text:
>MPSS_1
AAAAATGTCCGCCTGAAGAG
Or Upload a file containing one or more FASTA sequences
Browse...
Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)
Retrieve
Or Enter an existing ticket ID:
Retrieve
dna queries
peptide queries
Select the databases to search against
Select species:
Gallus_gallus
Gasterosteus_aculeatus
Homo_sapiens
dna database
Genomic sequence
peptide database
Known Consensus CDS Peptides (CCDS peptides)
Select the Search Tool
BLASTN
SSAHA2
TBLASTX
configure
RUN
Search sensitivity:
Near-exact matches (oligo)
Exact matches
Near-exact matches
Near-exact matches (oligo)
Allow some local missmatch
Distant homologies
No optimisation
About BlastView
Summary
setup
Not yet initialised
configure
Not yet initialised
results
Not yet initialised
display
Not yet initialised

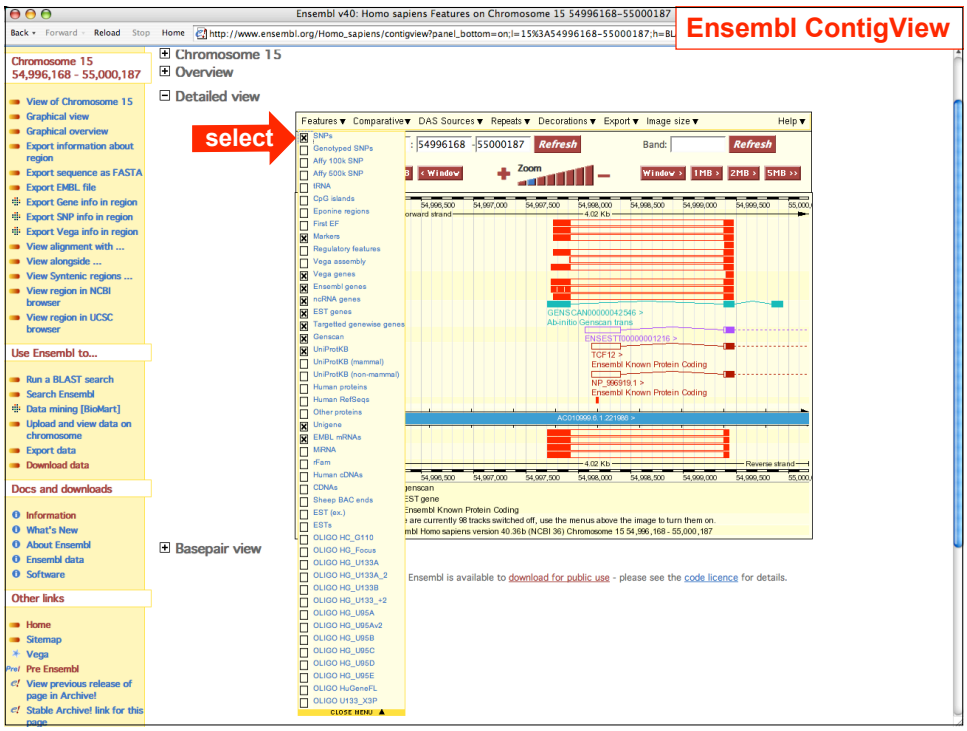

Ensembl BLAST search
BlastView
http://www.ensembl.org/Multi/blastview/BLA_bp8fHVoSi
Alignment Locations vs. Karyotype (click arrow to hide)
Alignment Locations vs. Query (click arrow to hide)
+ hsps
coverage
>MPSS_1
Alignment Summary (click arrow to hide)
Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)
refresh
click
100% identity over
100% of the query
length
Genomic sequence
BLASTN
Oligo sensitivity
configure
-E: 10
-B: 100
-filter: none
-RepeatMasker
-W: 11
-M: 1
-N: -3
-Q: 3
-R: 3
results
display
Not yet initialised

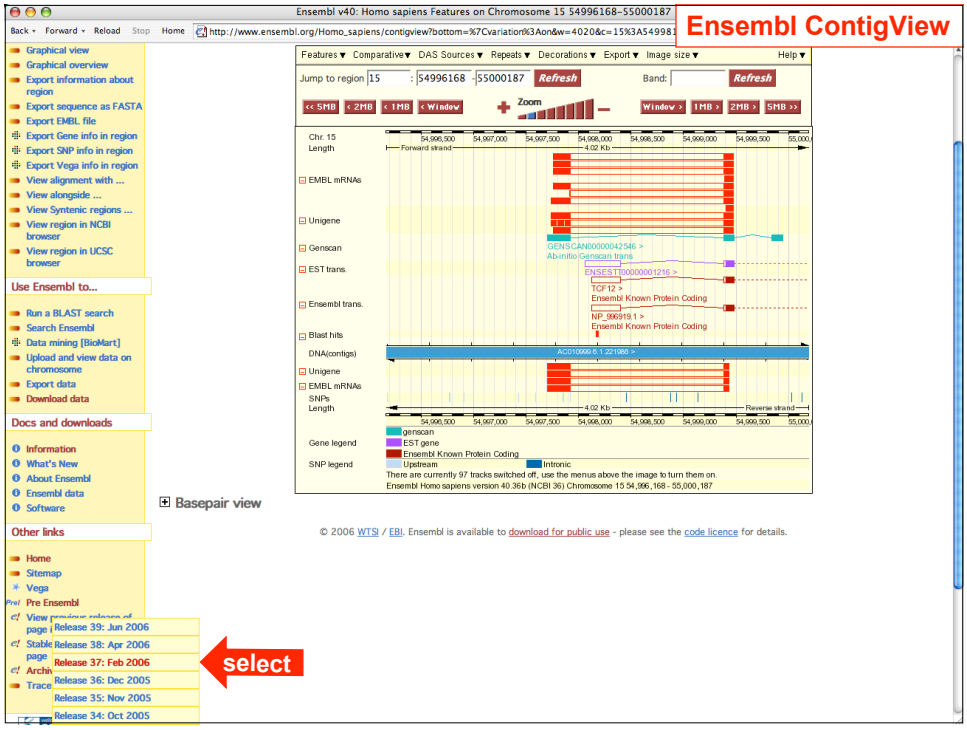

Ensembl v40: Homo sapiens Features on Chromosome 15 54996168–55000187
Ensembl ContigView
Ensembl Human ContigView
Chromosome 15
Overview
Detailed view
Basepair view
Chromosome 15 54,996,168 - 55,000,187
View of Chromosome 15
Graphical view
Graphical overview
Export information about region
Export sequence as FASTA
Export EMBL file
Export Gene info in region
Export SNP info in region
Export Vega info in region
View alignment with ...
View alongside ...
View Syntenic regions ...
View region in NCBI browser
View region in UCSC browser
Use Ensembl to...
Run a BLAST search
Search Ensembl
Data mining [BioMart]
Upload and view data on chromosome
Export data
Download data
Docs and downloads
Information
What's New
About Ensembl
Ensembl data
Software
Other links
Home
Sitemap
Vega
© 2006 WTSI / EBI. Ensembl is available to download for public use - please see the code licence for details.

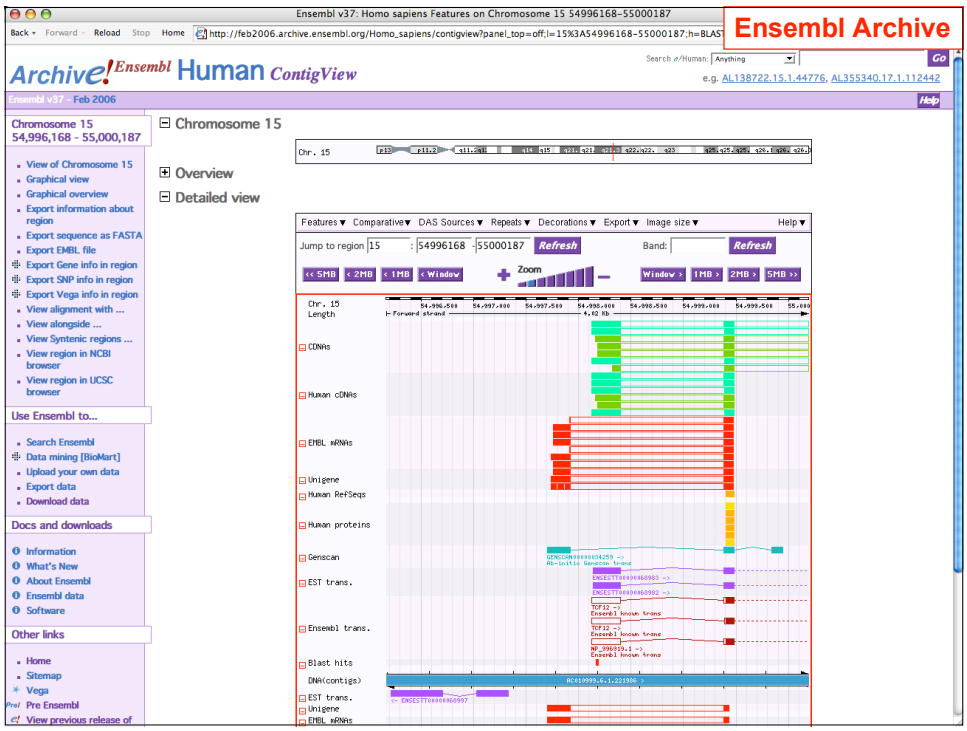

Ensembl v40: Homo sapiens Features on Chromosome 15 54996168–55000187
Ensembl ContigView
Ensembl Human ContigView
Chromosome 15
Overview
Detailed view
Basepair view
Features Comparative DAS Sources Repeats Decorations Export Image size Help
Jump to region 15 : 54996168 - 55000187 Refresh
Band: Refresh
Zoom
Chr. 15 Length
Forward strand
EMBL mRNAs
Unigene
Genscan
EST trans.
Ensembl trans.
Blast hits
DNA(contigs)
Reverse strand
Gene legend
genscan
EST gene
Ensembl Known Protein Coding
Chromosome 15 54,996,168 - 55,000,187
View of Chromosome 15
Graphical view
Graphical overview
Export information about region
Export sequence as FASTA
Export EMBL file
Export Gene info in region
Export SNP info in region
Export Vega info in region
View alignment with ...
View alongside ...
View Syntenic regions ...
View region in NCBI browser
View region in UCSC browser
Use Ensembl to...
Run a BLAST search
Search Ensembl
Data mining [BioMart]
Upload and view data on chromosome
Export data
Download data
Docs and downloads
Information
What's New
About Ensembl
Ensembl data
Software
Other links
Home
Sitemap
Vega
© 2006 WTSI / EBI. Ensembl is available to download for public use - please see the code licence for details.


Ensembl ContigView
Ensembl v40: Homo sapiens Features on Chromosome 15 54996168–55000187
Basepair view
Graphical view
Graphical overview
Export information about region
Export sequence as FASTA
Export EMBL file
Export Gene info in region
Export SNP info in region
Export Vega info in region
View alignment with ...
View alongside ...
View Syntenic regions ...
View region in NCBI browser
View region in UCSC browser
Use Ensembl to...
Run a BLAST search
Search Ensembl
Data mining [BioMart]
Upload and view data on chromosome
Export data
Download data
Docs and downloads
Information
What's New
About Ensembl
Ensembl data
Software
Other links
Home
Sitemap
Vega
Pre Ensembl
View previous release of page in Archive!
Stable Archive! link for this page
Archive! sites
Trace server
Genscan
EST trans.
Ensembl trans.
Blast hits
Amino acids
Sequence
DNA(contigs)


Ensembl
Add features to the ContigView

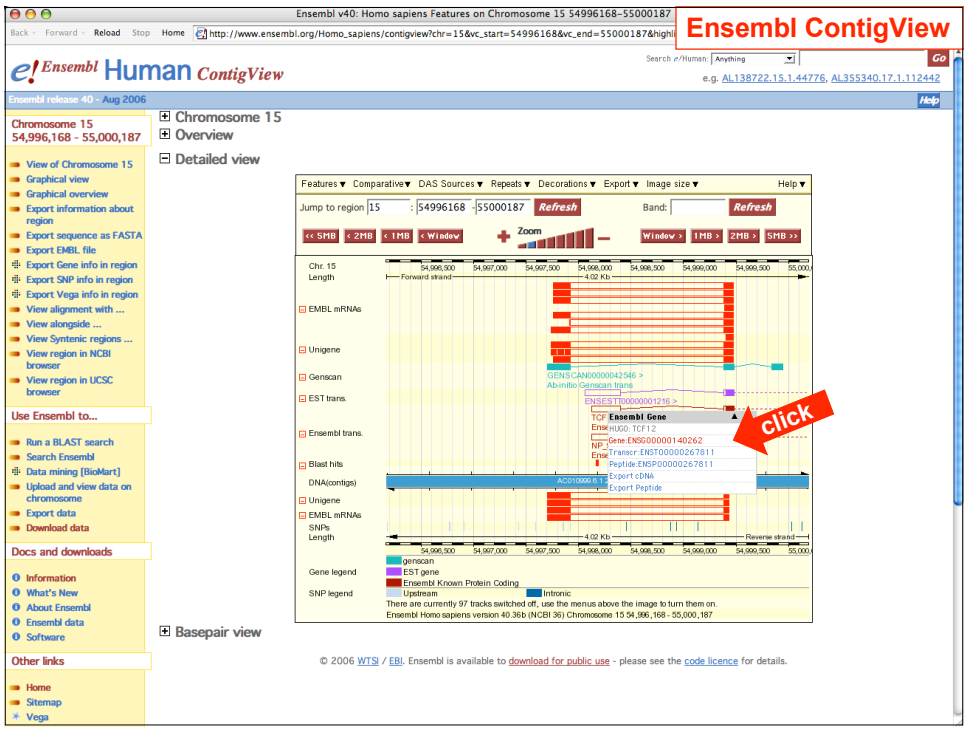

Ensembl ContigView
Ensembl v40: Homo sapiens Features on Chromosome 15 54996168–55000187
Chromosome 15
54,996,168 - 55,000,187
Chromosome 15
Overview
Detailed view
select
Basepair view
Ensembl is available to download for public use - please see the code licence for details.

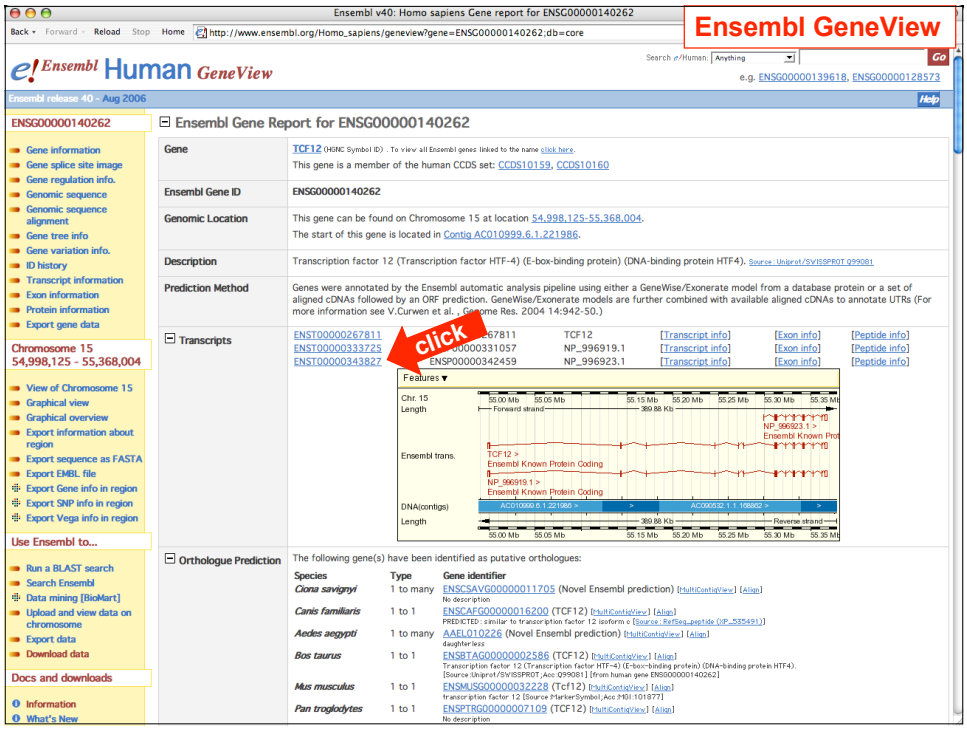

Ensembl ContigView
Ensembl v40: Homo sapiens Features on Chromosome 15 54996168–55000187
Basepair view
Release 39: Jun 2006
Release 38: Apr 2006
Release 37: Feb 2006
Release 36: Dec 2005
Release 35: Nov 2005
Release 34: Oct 2005
select
© 2006 WTSI / EBI. Ensembl is available to download for public use - please see the code licence for details.

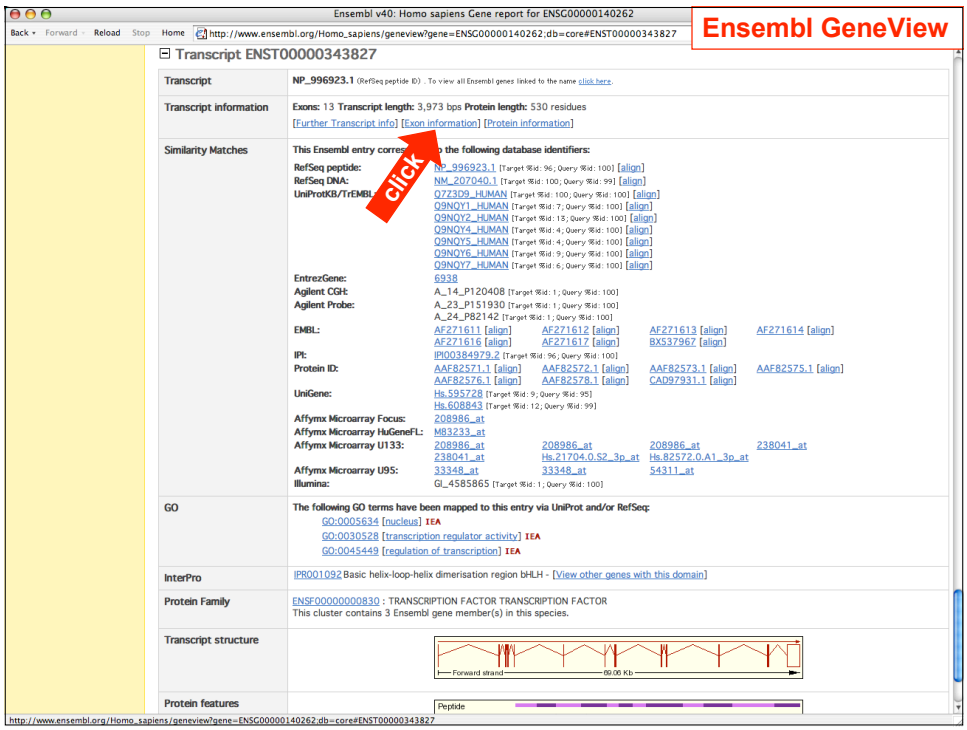


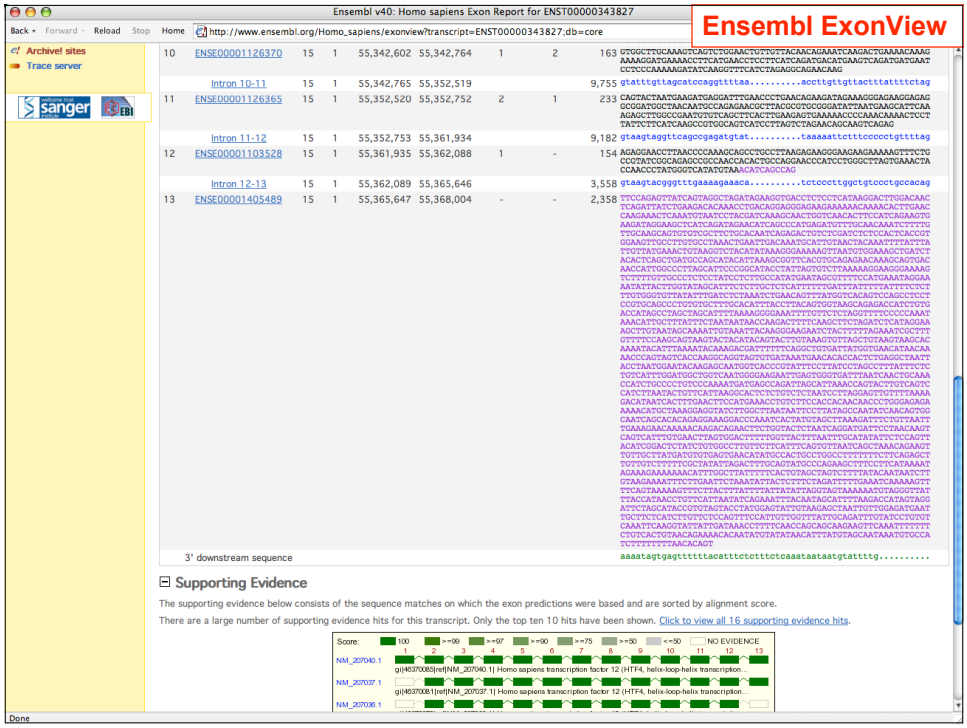
Ensembl

Get additional information about the gene, transcripts, and exons


Ensembl ContigView
Ensembl v40: Homo sapiens Features on Chromosome 15 54996168–55000187
Ensembl Human ContigView
Chromosome 15
Overview
Detailed view
Basepair view
click

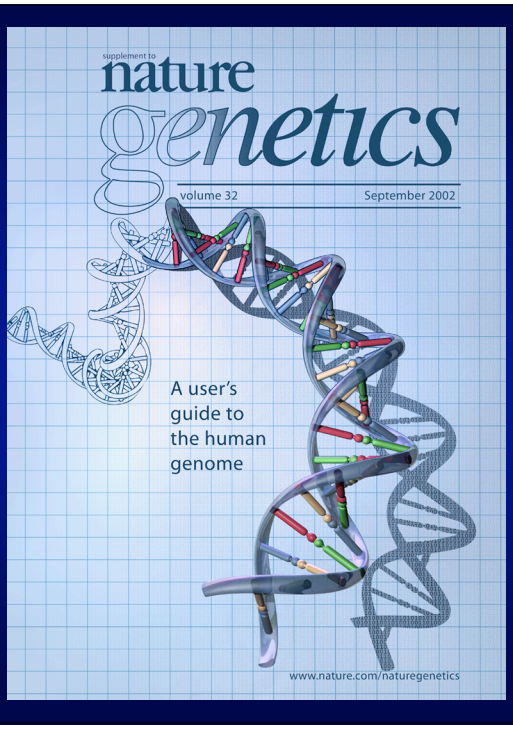

Ensembl GeneView
Ensembl v40: Homo sapiens Gene report for ENSG00000140262
Ensembl Human GeneView
Ensembl Gene Report for ENSG00000140262
Transcripts
Orthologue Prediction
click


Ensembl GeneView
Ensembl v40: Homo sapiens Gene report for ENSG00000140262
Transcript ENST00000343827
click


Ensembl ExonView
Ensembl v40: Homo sapiens Exon Report for ENST00000343827
Supporting Evidence

Additional resources

- UCSC Human Genome Browser User Guide
  http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html
- NCBI Genomic Biology
  http://www.ncbi.nih.gov/Genomes/
- NCBI MapViewer Help
  http://www.ncbi.nlm.nih.gov/mapview/static/MapViewerHelp.html
- Ensembl Worked Example
  http://www.ensembl.org/info/worked_example.pdf



http://www.nature.com/ng/supplements/

References

- Current Protocols in Bioinformatics
  UNIT 1.4: The UCSC Genome Browser
  UNIT 1.5: Using the NCBI Map Viewer to Browse Genomic Sequence Data
  Access through *http://nihlibrary.nih.gov/ResearchTools/OnlineJournals.htm*

- UCSC
  Hsu F *et al*. The UCSC Known Genes. *Bioinformatics*. 2006. 1034-46.
  Hinrichs AS *et al*. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* 2006. 34:D590-8.
  Kent WJ *et al*. Exploring relationships and mining data with the UCSC Gene Sorter. *Genome Res*. 2005. 15:737-41.
  Hsu F *et al*. The UCSC Proteome Browser. *Nucleic Acids Res*. 2005. 33:D454-8.
  Karolchik D *et al*. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*. 2004. 32:D493-6.
  Karolchik D *et al*. The UCSC Genome Browser Database. *Nucleic Acids Res*. 2003. 31:51-4.

- NCBI
  Wheeler DL *et al*. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2006:D173-80.
  Dombrowski SM and Maglott D. Using the Map Viewer to Explore Genomes. *in* The NCBI Handbook. 2003. *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books*

- Ensembl
  Birney E *et al*. Ensembl 2006. *Nucleic Acids Res.* 2006. 34:D555-61.
  Hammond MP, and Birney E. Genome information resources - developments at Ensembl. 2004. *Trends Genet.* 20:268-72.
  Birney E *et al*. An overview of Ensembl. 2004. *Genome Res*. 14:925-8.